CORRECTED VERSION*

IZATION

INTI ENT COOPERATION TREATY (PCT)

(51) Inte

C07H 21/04, C12N 1/20, ~~15/00~~ C07K 13/00

A1

Publication Number: WO 89/09223

(43) International Publication Date: 5 October 1989 (05.10.89)

(21) International Application Number: PCT/US89/01238

(22) International Filing Date: 24 March 1989 (24.03.89)

(31) Priority Application Numbers: 176,107 253,807 312,763

(32) Priority Dates: 30 March 1988 (30.03.88) 5 October 1988 (05.10.88) 21 February 1989 (21.02.89)

(33) Priority Country: US

(71) Applicant: ARCH DEVELOPMENT CORPORATION [US/US]; 1115-25 East 58th Street, Chicago, IL 60637 (US).

(72) Inventors: LIAO, Shutsung ; 5632 South Woodlawn Avenue, Chicago, IL 60637 (US). CHANG, Chawnshang ; 5644 South Harper Avenue, Chicago, IL 60637 (US).

(74) Agent: MEINERT, M., C.; Marshall, O'Toole, Gerstein, Murray & Bicknell, Two First National Plaza, Chicago, IL 60603 (US).

(81) Designated States: AT (European patent), BE (European patent), CH (European patent), DE (European patent), FR (European patent), GB (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent).

Published

With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: DNA BINDING PROTEINS INCLUDING ANDROGEN RECEPTOR

(57) Abstract

Disclosed are DNA sequences encoding DNA binding polypeptides including androgen receptor (AR) and TR2 polypeptides. Illustratively, human and rat AR-cDNA have 79 kD and 98 kD polypeptide expression products which are immunoprecipitable by human auto-immune anti-androgen receptor antibodies and are capable of binding androgens specifically and with high affinity. Also disclosed are antibodies and immunological methods and materials for detection of androgen receptor and TR2 polypeptides and hybridization methods and materials for detection of AR and TR2-related nucleic acids.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FR	France	ML	Mali
AU	Australia	GA	Gabon	MR	Mauritania
BB	Barbados	GB	United Kingdom	MW	Malawi
BE	Belgium	HU	Hungary	NL	Netherlands
BG	Bulgaria	IT	Italy	NO	Norway
BJ	Benin	JP	Japan	RO	Romania
BR	Brazil	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic			SE	Sweden
CG	Congo	KR	Republic of Korea	SN	Senegal
CH	Switzerland	LI	Liechtenstein	SU	Soviet Union
CM	Cameroon	LK	Sri Lanka	TD	Chad
DE	Germany, Federal Republic of	LU	Luxembourg	TG	Togo
DK	Denmark	MC	Monaco	US	United States of America
FI	Finland	MG	Madagascar		

"DNA BINDING PROTEINS INCLUDING ANDROGEN RECEPTOR"

BACKGROUND OF THE INVENTION

The present invention relates generally to DNA binding regulatory proteins and more particularly to DNA sequences encoding androgen receptor protein and novel DNA binding proteins designated TR2, to the polypeptide products of recombinant expression of these DNA sequences, to peptides whose sequences are based on amino acid sequences deduced from these DNA sequences, to antibodies specific for such proteins and peptides, and to procedures for detection and quantification of such proteins and nucleic acids related thereto.

There are five major classes of steroid hormones: progestins, glucocorticoids, mineralocorticoids, androgens, and estrogens. Receptor proteins, each specific for a steroid hormone, are distributed in a tissue specific fashion and in target cells, steroid hormones can form specific complexes with corresponding intracellular receptors. [Jensen, et al., Proc. Nat'l. Acad. Sci. (USA), 59:632 (1968); Gorski, et al., Ann. Rev. Physiol., 38:425-450 (1976); and Liao, et al., page 633 in Biochemistry of Hormones, H.L.J. Makin, ed. (Blackwell Sci. Publ. Oxford, 1984)]. The hormonal regulation of gene expression appears to involve interaction of steroid receptor complexes with certain segments of genomes and modulation of specific gene trans-

cription. See, e.g., Ringold, Ann. Rev. Pharmacol. Toxicol., 25:529 (1985); and Yamamoto, Ann. Rev. Genet., 19:209 (1985). Many of the primary effects of hormones involve increased transcription of a subset of genes in specific cell types.

The successful cloning of e.g., cDNAs coding for various steroid receptors has allowed the structural and functional analysis of different steroid receptor domains involved in steroid and DNA binding. See, e.g., Hollenberg, et al., Nature (London), 318:635 (1985); Miesfeld, et al., Cell, 46:389 (1986); Danielsen, et al., EMBO J., 5:2513 (1986); Greene, et al., Science, 231:1150 (1986); Green, et al., Nature (London), 320:134 (1986); Krust, et al., EMBO J., 5:891 (1986); Loosfelt, et al., Proc. Nat'l. Acad. Sci. (USA), 83:9045 (1986); Conneely, et al., Science, 233:767 (1987); Law, et al., Proc. Nat'l. Acad. Sci. (USA), 84:2877 (1987); Misrahi, et al., Biochem. Biophys. Res. Commun., 143:740 (1987); Arriza, et al., Science, 237:268 (1987); Sap, et al., Nature (London), 324:635 (1986); Weinberger, et al., Nature (London), 318:641 (1986); Benbrook, et al., Science, 238:788 (1987); and Evans, Science, 240:889 (1988).

Androgens, such as testosterone, are responsible for the development of male secondary sex characteristics and are synthesized primarily in testis. Cloning of a cDNA for androgen receptor (AR) has been difficult because, until recently, monospecific antibodies against AR have not been available for screening cDNA libraries. An abstract by Govindan, et al., J. Endocrinol. Invest., 10 (Suppl. 2) (1987), reported the isolation of cDNA clones encoding human androgen receptor isolated from a human testis λgt-11 cDNA library using synthetic oligonucleotides homologous to human glucocorticoid, estradiol, and progesterone receptors as probes. The expressed protein reportedly

bound tritium-labelled DHT (dihydrotestosterone) with high affinity and specificity. However, no nucleotide or amino acid sequence analysis was provided for full length androgen receptors, nor was any description provided concerning isolation of the full length putative androgen receptor clones.

Recently, Chang, C., et al., Science, 240:324 (April 15, 1988), co-authored by the inventors herein, described cDNAs encoding androgen receptors obtained from human testis and rat ventral prostrate cDNA libraries. These cDNAs for human and rat androgen receptor were reported to be long enough to code for 94 kDa and 76 kDa receptors. The molecular weights were derived with the assistance of a software program known as: DNA Inspector II (Textco West Lebanon, New Hampshire) open reading frame analysis. With a new DNA Inspector IIe program, hAR (918 amino acids) has an estimated molecular weight 98,608 and rAR (902 amino acids) has a molecular weight of 98,133. Therefore, the reported "94 kDa" AR is now termed "98 kDa" AR; and the hAR or rAR polypeptides, from the second ATG/Met, reported as "76 kDa" are now termed "79 kDa". See also, Chang, C., et al., Proc. Nat'l. Acad. Sci. (USA), 85:7211 (October 5, 1988) also co-authored by the inventors herein.

In contrast, Lubahn, D., et al., Science, 240:327 (1988), using libraries from human epididymis and cultured human foreskin fibroblasts obtained a human cDNA which was expressed in monkey kidney (COS) cells to yield a protein, present in the cytosol, capable of binding androgens. This cDNA, however, was only sufficient to code for a receptor having an estimated molecular weight of 41,000. Therefore, the cDNA obtained only coded for a portion of AR.

Of interest to the present invention is Young, et al., Endocrinol., 123:601 (1988), wherein the production of anti-AR monoclonal antibodies was reported. Anti-AR autoantibodies were identified in the sera of prostate cancer patients, as described in Liao, S., et al., Proc. Nat'l. Acad. Sci. (USA), 82:8345 (1984) (one of the co-inventors herein), and were characterized with respect to their titer, affinity, and specificity. Subsequently, lymphocytes from the blood of those patients having high antibody titers were isolated, transformed with Epstein-Barr Virus (EBV), and cloned for anti-AR monoclonal antibody production. These monoclonal antibodies were found to interact with androgen receptors from rat prostate. An attempt to scale-up antibody production resulted in a decline of antibody secretion. It is not uncommon for transformed B-cells to be more unstable than hybridoma cells. Kozbor, et al., Eur. J. Immunol., 14, 23 (1984). Because of the instability associated with such cell lines, an alternate source of monoclonal antibodies is preferred.

There thus exists a need in the art for information concerning the primary structural conformation of androgen receptor protein and other DNA binding proteins such as might be provided by knowledge of human and other mammalian DNA sequences encoding the same. Availability of such DNA sequences would make possible the application of recombinant methods to the large scale production of the proteins in procaryotic and eukaryotic host cells, as well as DNA-DNA, DNA-RNA, and RNA-RNA, hybridization procedures for the detection, quantification and/or isolation of nucleic acids associated with the proteins. Possession of androgen receptor and related DNA-binding proteins and/or knowledge of the amino acid sequences of the same would make possible, in turn, the development of monoclonal and

polyclonal antibodies thereto (including antibodies to protein fragments or synthetic peptides modeled thereon) for the use in immunological methods for the detection and quantification of the proteins in fluid and tissue samples as well as for tissue specific delivery of substances such as labels and therapeutic agents to cells expressing the proteins.

BRIEF SUMMARY OF THE INVENTION

The present invention provides novel purified and isolated DNA sequences encoding androgen receptor protein and a structurally related protein, designated TR2 protein, which also has DNA binding (and hence DNA replication or transcription regulatory) capacity. In presently preferred forms, novel DNA sequences comprise cDNA sequences encoding human and rat androgen receptor and human TR2 protein. Alternate DNA forms such as genomic DNA, and DNA prepared by partial or total chemical synthesis from nucleotides as well as DNA with deletions or mutations, is also within the contemplation of the invention.

Association of DNA sequences provided by the invention with homologous or heterologous species expression control DNA sequences, such as promoters, operators, regulators and the like, allows for *in vivo* and *in vitro* transcription to form messenger RNA which, in turn, is susceptible to translation to provide androgen receptor and TR2 proteins, and related poly- and oligo-peptides in large quantities. In a presently preferred DNA expression system of the invention, AR and TR2 encoding DNA is operatively associated with a viral (T7) regulatory (promoter) DNA sequence allowing for *in vitro* transcription and translation in a cell free system to provide, e.g., a 79 kD and 98 kD human androgen receptor (hAR) protein, 79 kD and 98 kD rat

androgen receptor (rAR) protein and smaller forms of these proteins, as well as TR2 protein, including 20 kD and 52 kD species.

Incorporation of DNA sequences into procaryotic and eucaryotic host cells by standard transformation and transfection processes, potentially involving suitable viral and circular DNA plasmid vectors, is also within the contemplation of the invention and is expected to provide useful proteins in quantities heretofore unavailable from natural sources. Systems provided by the invention included transformed E. coli DH5α cells, deposited January 25, 1989, with the American Type Culture Collection, 12301 Parklawn Drive, Rockville, Maryland 20852 in accordance with the U.S. Patent and Trademark Office's requirements for microorganism deposits, and designated EC-hAR3600 under A.T.C.C. Accession No. 67879; EC-rAR 2830, A.T.C.C. No. 67878; EC-TR2-5, A.T.C.C. No. 67877; and EC TR2-7, A.T.C.C. No. 67876. Use of mammalian host cells is expected to provide for such post-translational modifications (e.g., truncation, glycosylation, and tyrosine, serine or threonine phosphorylation) as may be needed to confer optimal biological activity on recombinant expression products of the invention.

Novel protein products of the invention include polypeptides having the primary structural conformation (i.e., amino acid sequence) of AR and TR2 proteins as well as peptide fragments thereof and synthetic peptides assembled to be duplicative of amino acid sequences thereof. Proteins, protein fragments, and synthetic peptides of the invention are projected to have numerous uses including therapeutic, diagnostic and prognostic uses and will provide the basis for preparation of monoclonal and polyclonal antibodies specifically immunoreactive with AR and TR2 proteins. Preferred protein fragments and synthetic peptides

include those duplicating regions of AR and TR2 proteins which are not involved in DNA binding functions and the most preferred are those which share at least one antigenic epitope with AR and TR2 proteins.

Also provided by the present invention are polyclonal and monoclonal antibodies characterized by their ability to bind with high immunospecificity to AR and TR2 proteins and to their fragments and peptides, recognizing unique epitopes which are not common to other proteins especially DNA binding proteins.

Illustratively provided according to the present invention are monoclonal antibodies, designated AN1-6, AN1-7, AN1-15; and produced by hybridoma cell lines designated H-AN1-6, H-AN1-7, H-AN1-15; deposited January 25, 1989, under Accession Nos. HB 10,000; HB 9,999; and HB 10,001, respectively, with the American Type Culture Collection, 12301 Parklawn Drive, Rockville, Maryland 20852 in accordance with the U.S. Patent and Trademark Office's requirements for microorganism deposits. These antibodies are characterized by (a) capacity to bind androgen receptors from rat ventral prostate and synthetic peptides having sequences predicted from the structure of hAR-cDNA and rAR-cDNA; (b) specific immunological reactivity with, and capacity to reversibly immunobind to, naturally occurring and recombinant androgen receptors, in native and denatured conformations; and (c) specific immunological reactivity with, and capacity to reversibly immunobind to, proteinaceous materials including all or a substantially, immunologically significant, part of an amino acid sequence duplicative of that extant at residues 331 through 577 of hAR and corresponding amino acid sequences in rAR.

The monoclonal antibodies of the invention can be used for affinity purification of AR from human or rat prostate, and other sources such as AR-rich organs and cultured cells.

Also provided by the present invention are novel procedures for the detection and/or quantification of normal, abnormal, or mutated forms of AR and TR2, as well as nucleic acids (e.g., DNA and mRNA) associated therewith. Illustratively, antibodies of the invention may be employed in known immunological procedures for quantitative detection of AR and TR2 proteins in fluid and tissue samples, of DNA sequences of the invention (particularly those having sequences encoding DNA binding proteins) that may be suitably labelled and employed for quantitative detection of mRNA encoding these proteins.

Among the multiple aspects of the present invention, therefore, is the provision of (a) novel AR and TR2-encoding DNA sequences set out in Figure 3, as well as (b) AR and TR2-encoding DNA sequences which hybridize thereto under hybridization conditions of the stringency equal to or greater than the conditions described herein and employed in the initial isolation of cDNAs of the invention, and (c) DNA sequences encoding the same allelic variant, or analog AR and TR2 polypeptides through use of, at least in part, degenerate codons. Correspondingly provided are viral or circular plasmid DNA vectors incorporating such DNA sequences and procaryotic and eucaryotic host cells transformed or transfected with such DNA sequences and vectors as well as novel methods for the recombinant production of AR and TR2 proteins through cultured growth of such hosts and isolation of these proteins from the hosts or their culture media.

Preferred polypeptide products of the invention include the approximately 79 kD (starting from the second ATG/Met) and 98 kD (starting from the first ATG/Met) hAR polypeptides having the deduced amino acid sequence of 734 and 918 residues, respectively, as set out in Figure 3. Also preferred are the 79 kD and 98 kD

rAR species polypeptides having the deduced sequence of 733 and 902 residues set out in Figure 3 and the 20 kD and 52 kD species human TR2 polypeptides having the same deduced amino acid sequence of 184 and 483 residues set out in Figure 4. The preferred 79 kD and 98 kD hAR and rAR polypeptides may be produced in vitro and are characterized by a capacity to specifically bind androgens with high specificity and by their immunoprecipitatability by human auto-immune anti-androgen receptor antibodies.

Other aspects and advantages of the present invention will be apparent upon consideration of the following detailed description thereof which includes numerous illustrative examples of the practice of the invention, reference being made to the drawing wherein:

Figure 1 illustrates the strategy employed in construction of a human androgen receptor cDNA vector;

Figure 2 illustrates the strategy employed in construction of rat androgen receptor cDNA vectors;

Figure 3 provides a 3715 base pair nucleotide sequence for a human androgen receptor (hAR) DNA clone and the deduced sequence of 734 and 918 amino acid residues for hAR proteins; and in addition provides a 3218 base pair nucleotide sequence for a rat androgen receptor (rAR) DNA clone and the deduced sequences of 733 and 902 amino acids for two rAR species;

Figure 4 provides a 2029 base pair nucleotide sequence for a human TR2 DNA clone and a deduced sequence of 483 amino acids for a "TR2-5" species and a deduced sequence of 184 amino acids for a "TR2-7" species; and

Figure 5 provides an amino acid sequence alignment of the cysteine-rich DNA binding domain of human androgen receptor, glucocorticoid receptor, mineralocorticoid receptor, progesterone receptor, estrogen receptor, TR2, rat AR, chick vitamin D receptor

(c-VDR), and the v-erb A oncogene product of avian erythroblastosis virus.

Figures 6, 7, and 8 illustrate, respectively, the in-frame fusion of three different parts of the AR gene (the N-terminal, the DNA-binding domain and the androgen-binding domain) to the N-terminal half of the trpE gene using pATH expression vectors.

DETAILED DESCRIPTION

The following examples illustrate practice of the invention. Example 1 relates to the isolation, preparation, and partial structural analysis of cDNA for human and rat androgen receptors. Example 2 relates to confirmation of the presence on the human X-chromosome of an AR-type cDNA sequence. Example 3 relates to the preparation of human and rat cDNAs containing AR-type cDNA from different clones and ligation into the pGEM-3Z plasmid. Example 4 relates to transcription and translation of the AR-type cDNA plasmid DNA. Example 5 relates to steroid binding activity of the expression product of Example 4. Example 6 relates to the binding activity of the expression product of Example 4 to human auto-antibodies. Example 7 relates to the characterization of TR2-cDNA. Example 8 relates to the *in vitro* transcription and translation of TR2-cDNA. Example 9 relates to the binding activity of TR2-cDNA expression product. Example 10 relate to the androgen regulation of TR2 mRNA levels in the rat ventral prostate. Example 11 relates to recombinant expression systems of the invention. Example 12 relates to the production of fusion proteins and their use in producing polyclonal and monoclonal antibodies according to the invention. Example 13 relates to use of DNA probes of the inventions. Example 14 relates to development of transgenic animals by means of DNA sequences of the invention.

These examples are for illustrative purposes only and are not intended in any way to limit the scope of the invention.

EXAMPLE 1

Preparation and Partial Structural Analysis of cDNA for Human and Rat Androgen Receptors

The isolation of cDNA for human androgen receptor (hAR) and rat androgen receptor (rAR) was accomplished using λGT11 cDNA libraries. The human testis and prostate λGT11 libraries were obtained from Clontech Co., Palo Alto, California and a rat ventral prostate λGT11 library in E.coli Y1090 was constructed as described in Chang, et al., J. Biol. Chem., 262:11901 (1987). In general, clones were differentiated using oligonucleotide probes specific for various steroid receptors. The cDNA libraries were initially screened with a set of 41-bp oligonucleotide probes designed for homology to nucleotide sequences in the DNA-binding domain of glucocorticoid receptors (GR), estrogen receptors (ER), progesterone receptors (PR), mineralocorticoid receptors (MR), and the v-erb A oncogene product of avian erythroblastosis virus. The set of probes had the following sequence: TGTGGAAGCTGT/CAAAGTC/ATTCTTAAAAGG/AGCAA/GTGGAAGG.

The plaques were replicated on a nitrocellulose filter and screened with a 5'-end ^{32}P-labeled 41-bp oligonucleotide probes. The conditions of hybridization were 25% formamide, 5X Denhardt's solution (0.1% Ficoll 400, 0.1% polyvinylpyrrolidone, 0.1% bovine serum albumin), 0.1% SDS, 5X SSC (1X SSC is 150 mM NaCl, 15 mM sodium citrate), 100 μg/ml denatured salmon sperm DNA, and 1 μg/ml poly(A) at 30°C. Filters were washed with a solution containing 0.1% SDS, 0.05% sodium pyrophosphate and 0.4X SSC at 37°C.

A less stringent hybridization condition (2X SSC at 37°C) was used for the first screen employing the 41 bp probes. The remaining clones were then probed again at more stringent conditions by reducing the concentration of SSC, eventually to 0.4X SSC at 37°C, or by increasing the temperature, or by increasing the concentration of formamide. In some procedures, 5X SSC, 8% dextran sulfate, and 20% formamide, at 42°C was employed and the result was equivalent to that obtained with 0.6X SSC.

From approximately 3 X 10^6 human testis recombinants and 6 X 10^5 rat ventral prostate recombinants, 302 and 21 positive clones, respectively, were obtained.

Based on the assumption that AR might have a cysteine-rich DNA binding domain highly homologous to the DNA-binding regions of other steroid receptors, positive clones from the first screenings were probed with 5'-end ^{32}P-labeled 24-bp oligonucleotides specific for the various steroid receptors for the possible presence of cDNA for AR through a process of elimination. The GR-cDNA clones were eliminated by screening with two GR-specific 24-bp probes that had nucleotide sequences identical to nucleotide segments immediately next to the 5'-end or the 3'-end of the DNA binding-region of hGR-cDNA , i.e., TGTAAGCTCTCCTCCATCCAGCTC and CAGCAGGCCACTACAGGAGTCTCA. 244 and 14 clones, respectively, were eliminated as hGR- and rGR-cDNA clones.

Using similar procedures involving four 24-bp probes for the 5'-end of PR(CCGGATTCAGAAA/GCCAGT/-CCAGAGC) and two 24-bp probes for the 3'-end of ER(GCA/-CGACCAGATGGTCAGTGCCTTG), no ER- or PR-cDNA clones were detected in the human testis library. In the rat prostate library, no ER-cDNA clones were detected but one positive clone was obtained with hPR-specific 24 bp probes.

Following this process of eliminating clones putatively encoding other steroid receptors, the DNA inserts in the remaining clones were analyzed by restriction mapping and subcloned into M13 vectors for di-deoxy sequence analysis. See, Chang, et al., J. Biol. Chem., 262:2826 (1987). Nucleotide sequence analysis allowed four clones to be identified as hMR-cDNA clones.

Through this stepwise process of elimination, 54 human testis clones and 6 rat prostate clones were selected and were then categorized into two groups: 30 human testis clones had sequences overlapping to form a 2.1 kb cDNA; and 24 human testis and 6 rat prostate clones had sequences overlapping to form a cDNA of about 2.7 kb. The two groups of cDNA were designated, respectively, as "TR2-type" and "AR-type" cDNA.

EXAMPLE 2

Confirmation of the Presence on the Human X-Chromosome of an AR-type cDNA Sequence Rather than a TR2-type cDNA Sequence

The length between the putative polyadenylation signal (AATAAA) and the 5'-end in the "TR-2 type" cDNA is only 2.0 kb, which is considerably shorter than that for the cDNA of other steroid receptors. Therefore, it was suspected that the "AR-type" cDNA, rather than the "TR2-type" cDNA, encoded androgen receptor. To obtain additional information, a human X-chromosome library prepared according to Kunkel, et al., Nucleic Acids Research, 11:7961 (1983) was probed with the TR2-type cDNA and AR-type cDNA of Example 1. With TR2-type cDNA fragments, no positive clones were detected, while 3 positive clones were obtained with a 1.9 kb fragment of AR-type cDNA from a human testis (clone AR 132), thereby confirming the presence of an AR-type cDNA

sequence on the human X-chromosome. Because the X-chromosome has been implicated as the chromosome which contains an AR gene [Lyon, et al., Nature (London), 227:1217 (1970); Meyer, et al., Proc. Nat'l. Acad. Sci. (USA), 72:1469 (1975); and Amrhein, et al., Proc. Nat'l. Acad. Sci. (USA), 73:891 (1976)], this information suggested that "AR-type" cDNA, but probably not the "TR2-type" cDNA, contained the DNA sequence that could encode for androgen receptor.

Two human clones containing DNA inserts that overlapped to form a 2.7 kb cDNA were designated AR 132 and AR 5. Two rat clones containing DNA inserts that overlapped to form a 2.8 kb cDNA were designated rAR 1 and rAR 4. After restriction enzyme digestion, the DNA segments from these AR-type clones were ligated, selected and amplified using pBR322 and pGEM-3Z vectors as described in Example 3 below.

EXAMPLE 3

A. Preparation of a Human cDNA Containing AR-type cDNA from Two Different Clones and Ligation Into the Cloning Vector pGEM-3Z Plasmid

Figure 1 relates to the strategy employed in the construction of a full length hAR-cDNA clone. cDNA of clone AR 132 was digested with Eco RI to obtain a 1.9 kb fragment which was then digested with Kpn I to provide a 1 kb Eco RI-Kpn I fragment. This 1 kb fragment was ligated to a 3 kb fragment obtained by digestion of clone AR 5 with Kpn I and Pvu I. The resulting 4 kb fragment was inserted into Eco RI and Pvu I-digested pBR322 vector and used to infect *E. coli* DH5α. The transformed clones were selected by tetracycline-resistance. The plasmid with the DNA insert was digested with Cla I and Nde I to obtain a 2.6 kb fragment. The fragment was blunt-ended with the Klenow fragment of *E.*

coli DNA polymerase I and ligated to the cloning vector pGEM-3Z plasmid DNA (Promega Biotec, Madison WI.) which was previously blunt-ended by digestion with Sma I. E. coli DH5 α cells were transformed with the plasmid so formed (designated plasmid PhAR3600) and colonies containing the plasmid were selected by ampicillin resistance and amplified. E. coli DH5α cells, transformed with plasmid PhAR3600, were designated EC-hAR3600 and were deposited with the American Type Culture Collection, 12301 Parklawn Drive, Rockville, Maryland 20852 on January 25, 1989 under Accession No. 67879.

The plasmid DNA was isolated and its structure analyzed by restriction enzyme mapping and sequencing. The 2.0 kb hAR fragment obtained by NruI-BamHI digestion of a 2.6 kb hAR in pGEM3Z was then ligated to another 1.6 kb ECORI-NruI fragment of hHR to obtain the full length 3715 bp hAR. The open reading frame is about 2.8 kb which is sufficient to code for a protein with more than 900 amino acids. Near the middle of the protein is a cysteine-rich region with a 72 amino acid sequence highly homologous to regions in other steroid receptors considered to be the DNA binding domain.

As set out in detail below and illustrated in Figure 2, formation of "full length" rat AR clones by slightly varying procedures results in constructions providing RNA transcripts translatable to 79 kD and 98 kD protein products.

B. Preparation of a Rat 2.7 kb cDNA and Ligation Into the Cloning Vector pGEM-3Z Plasmid

The 2.4 kb Eco RI-Eco RI cDNA insert of clone rAR 1 was digested with Xmn I to obtain a 2.3 k b fragment. This 2.3 kb Xmn I-EcoR I fragment was ligated to a 400 bp fragment that was obtained by

digestion of another cDNA clone insert (Eco RI-Eco RI insert of rAR 4) with Pst I. The ligated 2.7 kb fragment was inserted into Sma I and Pst I-digested pGEM-3Z vector and used to infect _E. coli_ DH5α. The _E. coli_ DH5α cells were transformed with the plasmid and colonies containing the plasmid were selected by ampicillin resistance and amplified. These cells were designated EC-rAR 2830 and were deposited with the American Type Culture Collection, 12301 Parklawn Drive, Rockville, Maryland 20852 on January 25, 1989 under Accession No. 67878. As noted in Figure 2, this construction allowed for a transcription product translated beginning with the second of two in-frame methionine-specifying codons (designated ATG_2).

C. Preparation of a Rat 2.83 kb cDNA Ligation Into the Cloning Vector pGEM-3Z Plasmid

The 2.4 kb Eco RI-Eco RI cDNA insert of rAR 1 was digested with Hind III to obtain a 1.68 kb fragment. The 1.68 kb Eco RI-Hind III fragment was ligated to a 1.15 kb DNA fragment obtained by digestion of another cDNA clone insert (rAR 6) with Hind III and Pst I. The ligated 2.83 kb fragment was inserted into Eco RI and Pst I-digested pGEM 3Z vector and used to infect _E. coli_ DH5α. _E. coli_ (DH5α) cells were transformed with the plasmid and colonies containing the plasmid were selected by ampicillin resistance and amplified. As noted in Figure 2, this construction allowed for a transcription product translated beginning at the first of two in-frame methionine-specifying codons (designated ATG_1).

Figure 3 provides the nucleotide sequence of the DNA sequence of the longer "full length" rat and human AR clones and includes the deduced amino acid sequences. The first and second methionine-specifying

codons are designated at amino acid positions 1 and 170 of rAR and positions 1 and 185 of hAR.

EXAMPLE 4

Transcription and Translation of the Human AR-type cDNA Plasmid in a Rabbit Reticulocyte Lysate System

pGEM-3Z vector (20 μg) containing 2.6 kb hAR DNA segment, as described in Example 3, was linearized with restriction enzyme Bam HI, phenol/chloroform extracted, and precipitated with ethanol. The linearized plasmid was transcribed in a reaction mixture containing 40 mM Tris-HCl, pH 7.5, 6 mM $MgCl_2$, 2 mM spermidine, 10 mM NaCl, 10 mM DTT, 500 μM each of ATP, GTP, CTP, and UTP, 160 units ribonuclease inhibitor, 5 μg plasmid, 30 units T7 RNA polymerase (Promega Biotec, Madison, WI) and diethylpyrocarbonate (DEPC)-treated water to a final volume of 100 μl. T7 RNA polymerase was used in the transcription of the plasmid DNA, because a T7 promotor, rather than the SP6 promotor, was found ahead of the 5'-end of the ligated AR-cDNA.

The reaction was allowed to proceed for 2 hrs. at 40°C. RQ1 DNase I (5 units) was added and the reaction continued for 15 mins. at 40°C. The reaction mixture was extracted with phenol/chloroform (1:1) and then with chloroform. RNA product was precipitated by the addition of 0.1 volume of 3 M Na-acetate and 2.5 volumes of ethanol, re-suspended in 0.5 M NaCl, and re-precipitated with 2.5 volumes of ethanol. RNA transcribed was isolated and then translated in a rabbit reticulocyte lysate system.

Translation of RNA was carried out in a micrococcal nuclease-treated rabbit reticulocyte lysate (Promega Biotec, Madison, WI) pre-mixed kit (100 μl) in the presence of 8 μg mRNA, 40 uCi of [^{35}S] methionine

(800 Ci/mmol; Amersham Co., Arlington Heights, IL) and 100 μM each of amino acid mixture without methionine. The reaction was allowed to proceed for 1 hour at 30°C. To quantitate the incorporation of radioactive methionine, 3 μl of the reaction mixture were added to 1 ml of 1 M NaOH containing 1.5% H_2O_2, 1 mM methionine, and 0.04% bovine serum albumin. The mixture was incubated for 15 mins. at 37°C to hydrolyze [^{35}S] methionine charged tRNA. The radioactive protein products were precipitated by the addition of 1 ml of 25% tricholoacetic acid and the radioactivity associated with the precipitates was determined.

By SDS-PAGE (8% acrylamide gel) analysis, performed as described in Saltzman, et al., J. Biol. Chem., 262:432 (1987), it was found that a 79 kD protein comprised more than 85% of the translated products.

EXAMPLE 5

Binding Activity of the 79 kD hAR Protein to a Synthetic Androgen

To study the steroid binding activity of the protein coded for by the cloned cDNA, the reticulocyte lysate of Example 4, containing the newly synthesized protein was incubated with 17α[^{3}H]-methyl-17β-hydroxy-estra-4,9,11-trien-3-one ([^{3}H] R1881), a potent synthetic androgen that binds AR with high affinity [Liao, et al., J. Biol. Chem., 248:6154 (1973)].

Specifically, RNA transcribed from the cloned cDNA, as described in Example 4, was translated in a rabbit reticulocyte lysate system and aliquots of the lysate were then incubated with 5 nM [^{3}H] R1881 (87 Ci/mmol) in the absence or presence of 25 nM, 50 nM, or 250 nM of non-radioactive steroid. The final incubation volume was 100 μl. The radioactive androgen binding was measured by the hydroxylapatite-filter method as des-

cribed in Liao, S., et al., J. Steroid Biochem., 20:11 (1984). The result was expressed as a percentage of the radioactivity bound in the control tube (5000 dpm) without additional non-radioactive steroid and is listed in Table 1.

TABLE 1

Androgen-specific binding of hAR coded by cloned cDNA

Non-radioactive steroid added	$[^3H]$ R1881-bound (% of control) 25 nM	50 nM	250 nM
R1881	13	10	1
5α-dihydrotestosterone	25	17	6
5β-dihydrotestosterone	89	89	81
17β-Estradiol	91	91	86
Progesterone	100	91	92
Dexamethasone	100	93	93
Hydrocortisone	96	90	90
Testosterone	38	28	Not tested

As shown in Table 1, the active natural androgen, 17β-hydroxy-5α-androstan-3-one(5α-dihydro-testosterone) competed well with $[^3H]$ R1881 binding, but the inactive 5β-isomer did not compete well with $[^3H]$ R1881 suggesting that it does not bind tightly to AR. The binding activity was steroid specific; dexamethasone, hydrocortisone, progesterone, and 17β-estradiol did not compete well with the radioactive androgen for binding to the 79 kD protein.

Similar steroid binding specificities have also been observed for rAR coded for by cloned cDNA. Chang, C., et al., Proc. Nat'l. Acad. Sci. (USA), 85:7211-7215 (1988).

Using the hydroxylapatite filter assay method, it was observed that approximately one molecule of the

^{35}S-labelled 79 kD protein obtained from the lysate bound about one molecule of the tritiated androgen at a saturating concentration of ligand. By Scatchard plot analysis, the apparent dissociation constant was 0.31 nM, which is similar to the binding constant (0.65 nM) reported previously for AR of rat ventral prostate as reported in Schilling, et al., The Prostate, 5:581 (1984).

EXAMPLE 6

Binding Activity of the 79 kD Protein to Human Auto-antibodies

It has previously been reported [Liao, et al., Proc. Nat'l. Acad. Sci. (USA), 82:8345 (1985)] that some older men with prostate cancers have high titers of auto-immune antibodies to AR in their serum samples. The ability of human auto-antibodies to recognize the 79 kD protein made by the reticulocyte lysate system was therefore studied. The receptor protein made in the lysate system of Example 4 was incubated with [^{3}H] R1881 to allow the formation of radioactive androgen-androgen receptor (A-AR) complexes and was then mixed with serum containing auto-antibodies.

Reticulocyte lysate containing translated AR was incubated with [^{3}H] R1881, as described in Example 4, and then incubated again in either the presence of or absence of 5 μl of human male serum containing anti-bodies to AR (anti-AR serum) for 4 hrs. at 4°C. Rabbit serum containing anti-human immuglobulins (Anti-IgG) was then added as the second antibody. After 18 hrs. of incubation at 4°C, the mixture was centrifuged and the radioactivity associated with the precipitate was esti-mated. Human female serum, not containing anti-AR anti-body, was also used for comparison.

The results shown in Table 2 below, indicate a quantitative immunoprecipitation of the radioactive A-AR complexes in the presence of both the high titer human serum and a rabbit anti-human immunoglobulin IgG. By SDS-PAGE, it was also observed that the immunoprecipitated protein was the 79 kD protein.

TABLE 2

Anti-human immunoglobulin-dependent precipitation of hAR made by the translation of RNA transcribed from cloned cDNA

Sample incubated with [^{3}H]R1881	Anti-serum addition	Immunoprecipitable radioactivity(dpm)
AR coded by cDNA[a]	None	32
	+Anti-AR serum + Anti-IgG	8212
	+Female serum + Anti-IgG	430
	+Anti-IgG	8
Heated AR[b]	+Anti-AR serum + Anti-IgG	42
BMW-lysate[c]	+Anti-AR serum + Anti-IgG	204

[a] 8500 dpm of the radioactive AR complexes made were used.

[b] Reticulocyte lysate containing AR was heated at 50°C for 20 mins. to inactivate receptor and release the radioactive androgen bound before the addition of antiserum.

[c] Brome Mosaic Virus RNA was used in the reticulocyte lysate translation system instead of RNA transcribed from cloned cDNA.

EXAMPLE 7

Characterization of TR2-cDNA

Of the more than 40 TR2-type human cDNA clones obtained, including the 30 described in Example 1, the clone designated TR2-5 was found to be 2029 base pairs in length as indicated in Figure 4. The open reading frame between the first ATG and terminator TAA can encode 483 amino acids with a calculated molecular weight of 52 kD. The putative DNA binding region is underscored. The putative initiator ATG matched closely with Kozak's concensus sequence for active start codons. [See, Kozak, M., Nature, 308:241 (1984).] Two triplets upstream of this ATG codon is an in-frame terminator (TAA) further supporting initiator function for the ATG. Eleven out of the 30 TR2-type clones of Example 1, as represented by the clone designated TR2-7, contain an internal 429 bp insertion between nucleotide sequence 669 and 670 (designated by an asterisk in Figure 4). This internal insertion introduces a termination codon TAG (underscored in the insert sequence footnote) which reduces the open reading frame to 184 amino acids with a calculated molecular weight of 20 kD. It is likely that the insertion in these 11 TR2 clones (or deletion in the 19 other TR2 clones) represents either the existence of two types of mRNA in the human testis or an artifact of cDNA construction. In the 3'-nontranslated region, a eukarotic polyadenylation signal AATAAA is present between the nucleotide sequence 2000 and 2007 of the TR2-5 clone.

Other variants of TR-2 with open reading frames at the putative ligand-binding domains have been obtained. Some of these may code for receptors for new hormones or cellular effectors. It is anticipated that the knowledge of TR2-cDNA sequences will be utilized in isolation and structural analysis of other cellular

receptors, their genes, and ligands (endogenous or therapeutic agents) that can regulate cellular growth and functions in both normal and diseased organs.

Figure 5 depicts an amino acid sequence alignment of the cysteine-rich DNA binding domain of human androgen receptor, glucocorticoid receptor, mineralocorticoid receptor, progesterone receptor, estrogen receptor, human TR2 protein, rat AR, chick vitamin D receptor (c-VDR), and the v-erb A oncogene product of avian erythroblastosis virus. The numbers in the left margin represent the positions of amino acid residues in the individual receptors. Common residues are boxed with solid lines. The residues in dotted boxed represent those not in common with those in the solid boxes. V-erb A has two more amino acids at the starred position.

In this region, the human and rat cDNAs for AR have identical amino acid sequences, although for some amino acids different codons are employed. Also in this region, the homology between human AR or rat AR and other receptors is as follows: glucocorticoid receptor (GR), 76.4%; mineralo-corticoid receptors (MR), 76.4%; progesterone receptors (PR), 79.2%; estrogen receptors (ER), 55.6%; TR2, 45.8%; chick vitamin D receptor (c-VDR), 40,3%; and the v-erb A oncogene product of avian erythroblastosis virus, 40.3%. In the putative region for steroid binding, which has about 200 amino acids near the -COOH terminal of steroid receptors, the homology between human AR or rat AR and hGR, hMR, or hPR is about 45-55%, whereas the homology between human AR and rat AR and hER is less than 20%. Thus, human and rat AR appear to be more closely related to GR, MR, and PR, than to v-erb A or to receptors for estrogen, vitamin D, and thyroid hormones.

The DNA binding domain of TR2 (amino acids 111 to 183) has a high homology with the steroid receptor

super-family as follows: retinoic acid receptor (RAR), [Giguere, et al., Nature, 330:624 (1987)], 65%; thyroid receptor (T_3R) [Sap, et al., Nature, 324:635 (1987)], 59%; mineralocorticoid receptor (MR), [Arriza, et al., Science, 235:268 (1987)], 54%; vitamin D_3 receptor (VD_3R) [McDonnell, et al., Science, 235:1214 (1987)], 53%; hERR1 and hEER2, [Giguere, V., et al., Nature, 331:91 (1988)], 51% estrogen receptor (ER), [Hollenberg, et al., Nature, 318:635 (1985)], 51%; glucocorticoid receptor (GR) [Hollenberg, et al., Nature, 318:635 (1985)], 50%; androgen receptor (AR), 50%; progesterone receptor (PR), 49%; [Loosfelt, et al., Proc. Nat'l. Acad. Sci.,(USA), 83:9045 (1986)]. As noted in Figure 5, the positions of 20 amino acids (9 Cys, 3 Arg, 2 Gly, 2 Phe, 1 Lys, 1 Met, 1 Asp, 1 His) in the putative DNA binding domain are identical among all isolated thyroid steroid receptor genes. It has been proposed that this highly conserved region may be involved in the formation of a DNA binding finger. See, Weinberger, et al., Nature, 318:670 (1985). Like the other steroid receptors, TR2 does not have the two extra amino acids (Lys-Asn) found only in the thyroid receptors' DNA binding domain. See, Sap, et al., Nature, 324:635 (1987).

EXAMPLE 8

In Vitro Transcription and Translation of TR2 cDNA

The Eco RI-Eco RI DNA inserts from clones TR2-5 and TR2-7 were isolated and ligated to an EcoR1 digested pGEM-3Z vector for *in vitro* transcription essentialy as described in Example 3. *E. coli* DH5α cells, transformed with these plasmids were designated EC TR2-5 and EC TR2-7 and were deposited with the American Type Culture Collection, 12301 Parklawn Drive,

Rockville, Maryland 20852 on January 25, 1989 under Accession Nos. 67877 and 67876.

Transcribed RNA was then translated in a rabbit reticulocyte lysate system. By SDS-polyacrylamide gel electrophoresis (PAGE), it was found that the major translated product of TR2-7, which has an internal 429 bp, insertion, was a 20 kD protein. The major translated product of TR2-5 was a 52 kD protein.

To further characterize these translated proteins, the translation lysate was passed over a DNA cellulose column. The bound product was then eluted, concentrated and applied to SDS-PAGE. The results indicated that the translated proteins were indeed DNA-binding proteins.

EXAMPLE 9

Binding Activity of TR2-5 cDNA Expression Product

To study the steroid binding activity of the translation products of the TR2-5 clone, the products were incubated with all major classes of steroids, including androgens, progesterone, glucocorticoid and estrogen but no significant binding with the above steroids was observed. This does not necessarily rule out a steroid binding function for the protein. Possibly the TR2-5 expression product steroid binding activity may involve some post-translation modifications missing in the rabbit reticulocyte lysate system. Alternatively, the TR2-5 translated protein may be steroid independent or may bind to an unidentified ligand present in the human testis or rat ventral prostate.

The size of TR2 mRNA was determined by Northern blot analysis with TR2-5 cDNA insert as a probe. One 2.5 kb band was detected which should

include enough sequence information to code for a 52 kD protein. The TR2 mRNA tissue distribution was also analyzed by dot hybridization. The hybridization was visualized by densitometric scanning of the autoradiographs, individual dots were cut and radioactivity measured by liquid scintillation counting [See, Chang, et al., J. Biol. Chem., 262:2826 (1987)]. The results showed that TR2 mRNA was most abundant in the rat ventral prostate with the relative amounts in other tissues being: prostate 100%, seminal vesicle 92%; testis, 42%; submaxillary gland, 18%; liver, 13%; kidney, <1%; and uterus, <1%.

EXAMPLE 10

Analysis of Androgen Regulation of AR and TR2 mRNA Levels in the Rat Ventral Prostate

Because rat ventral prostate is an androgen-sensitive organ and contains the greatest amount of AR and TR2 mRNA, the effect of androgen depletion and replacement on the mRNA levels was studied by RNA dot hybridization and Northern blot analysis. Total RNA was extracted from the ventral prostate of normal rats, rats castrated and rats previously castrated and treated with 5α-dihydrotestosterone (17β-hydroxy-5α-androstand-3-one). AR mRNA levels per unit of DNA increased 200 to 300% of the level for normal rats within 2 days after castration. Administration of 5α-dihydrotestosterone (5 mg/rat/day) into castrated rats reduced the AR mRNA level to that of normal rats. TR2 mRNA levels, per unit of DNA, were increased to 170% of the normal rat within 2 days after castration. Injection of 5α-dihydrostestosterone (5 mg/rat/day) into castrated rats reduced the TR2 mRNA to the levels of normal rats. Interestingly, the total prostate RNA levels, at the same period of time, were decreased to 40% of the normal level. The

effects of androgen on the levels of prostatic TR2 mRNA were further confirmed by flutamide injection experiments. Flutamide, an anti-androgen which antagonizes the effects of 5α-dihydrotestosterone on the ventral prostate weights in castrated rats [Neri, et al., Invest. Urol., 10:123 (1972)], was injected into normal rats for from 2 to 6 days. TR2 mRNA levels were then measured by dot hybridization as described above. The results show that flutamide injection, like castration, increased TR2 mRNA levels. The change in the AR or TR2 protein levels could be due to a change in mRNA stability and utilization or a change in the regulation of gene transcription. The activation or inactivation by androgen of specific genes to different degrees in the same organ may suggest that androgen is involved in the structuring of the pattern of gene expression in the target cell. Also, if androgen-mediated gene repression mechanisms are related to growth of the prostate, then a further study of the mechanism and structure of genes, repressed AR and TR2 mRNA may provide a better understanding of androgen action in the normal and abnormal prostate and other hormone sensitive organs.

Also, defects in the structures of AR and androgen sensitive genes and/or losses of the control of the production and function of these gene products can be the causes of the abnormal growth of androgen sensitive or insensitive tumors like prostate cancers. These lines of research may, therefore, be helpful in designing new diagnostic methods and treatments for patients.

EXAMPLE 11

Expression of Cloned AR-Genes and Androgen Sensitive Genes in Eukaryotic and Prokaryotic Cells

The ability of cloned genes to function when introduced into mammalian, yeast, and bacterial cells has proved to be very valuable in understanding the function and regulatory mechanism of genes. Recombinant techniques can provide, in large quantities, gene expression products (proteins) which are not readily obtainable from natural sources. While bacterial systems are very useful in large scale production of those proteins which do not require substantial post-translational modification for optimal biological activity, eukaryotic systems are particularly advantageous because of their ability to correctly modify the expressed proteins to their functional forms.

Using well known techniques, AR-cDNA and TR2-cDNA may readily be used for large scale production of gene products. For this purpose, the most efficient transcription units can be constructed using viral, as well as non-viral, vectors with regulatory signals that can function in a variety of host cells. SV40, pSV2, adenoviruses, and bovine papilloma virus DNA have been used successfully for introduction of many eukaryotic genes into eukaryotic cells and permit their expression in a controlled genetic environment. These and similar systems are expected to be appropriate for the expression of AR- and TR2-genes. To assist gene transfer, the two most widely used methods, the "calcium phosphate precipitation" and the "DEAE-dextran technique" can be used. Genes can be introduced into cells either transiently, where they continue to express for up to 3 days, or, more permanently to form stably transformed cell-lines. The expressed proteins can be detected by androgen binding or antibody assays.

The expression of cloned AR-genes was achieved as follows in a eukaryotic system. NIH 3T3 cells, contact-inhibited cells established from NIH Swiss mouse embryo, were co-transfected with hAR cDNA inserted into pBPVMTH vectors as described by Gorman, "DNA Cloning", 2:143-190 D. M. Glover, ed.; (Oxford, Washington, D.C. 1985). Transfected cells were cloned and grown in multiple-well cell culture plates. About 100 individual cell lines were isolated. Of these, 6 demonstrated [^{3}H] R1881-binding activity at least 4-fold the activity of cells transfected with pSV2 vector alone, i.e., without the hAR cDNA sequence.

To express AR cDNA in prokaryotic systems, hAR and rAR cDNAs were inserted into a number of expression vectors including pUR, λGT11, pKK223-3, pKK233-2, pLEX, pATH1, pATH2, pATH10, and pATH11. Vectors with AR cDNA inserts were used to infect *E. coli* strains (JM109, DH5α, Y1089, JM105, and RR1). According to polyacrylamide gel electrophoresis analysis, the infected bacteria can synthesize AR fragments coded for by the AR cDNA inserts. Some of these AR polypeptides are degraded in culture. Amino terminal, DNA-binding, and androgen binding domains were used, as described in Example 12, to construct fusion proteins representing these domains.

EXAMPLE 12

Production of Polyclonal and Monoclonal Antibodies to AR

The isolation of AR in significant amounts from androgen sensitive organs has been exceedingly difficult. Therefore, the high-level expression of hAR or rAR cDNAs, as shown in Example 11, is expected to be an ideal way for the large scale production of AR. In addition, oligopeptides, with sequences identical to the

deduced amino acid sequences of portions of AR molecules, can be chemically synthesized inexpensively in large quantities. Both AR produced by expression vectors in eukaryotic or prokaryotic cells and AR oligopeptides chemically synthesized were used as antigens for the production of monoclonal antibodies as described in greater detail below.

Generally, several chemically synthesized oligopeptides, representing sequences unique to AR, (i.e., PYGDMRLETARDHVLP; CPYGDMRLETARDHVLP; and SIRRNLVYSCRGSKDCIINK) were bound to BSA or KLH carrier proteins and were used to immunize mice. Spleen cells from these mice were fused to myeloma cells to produce hybrid antibody producing cells. Analysis by ELISA (enzyme-linked immunoassay) of the supernatants of 4 hybrid cultures appeared to indicate the presence of immunoglobulin that interacts with AR of rat ventral prostate. It is anticipated that these cells which produce monoclonal antibodies can be injected intraperitoneally into BALB/c mice previously treated with pristane. Ascites fluids can then be harvested and antibodies precipitated with ammonium sulfate.

Expression of Androgen Receptor Fusion Protein in E. coli

Three different parts of the AR gene (encompassing the N-terminal domain, the DNA-binding domain and the androgen-binding domain) were fused, in frame, to the N-terminal half of the trpE gene (trpE promoter-the first 969 bp of trpE coding region-multiple cloning region of pUC12) by using the pATH expression vectors as shown in Figures 6, 7, and 8, respectively. Dieckmann, et al., J. Biol. Chem., 260:1513 (1985).

These constructions resulted in the fusion of approximately 25 kDa of AR, including a portion of the N-terminal domain; 29 kDa of AR, including a major

portion of the DNA-binding domain; and 12 kDa of AR, including a portion of the androgen-binding domain; to the 33 kDa trpE protein. Because the trpE protein is insoluble, partially purified induced fusion proteins were obtained simply by lysing the E. coli and precipitating the insoluble fusion proteins. After electrophoresis on SDS-polyacrylamide gels, the induced fusion proteins, i.e., those proteins not present in the control pATH vector (no AR gene insert), were sliced from the gels and then used for immunization.

Fusion proteins, other than the three specifically exemplified, can also be constructed using these means.

Production and Purification of Anti-AR Antibodies

Rabbits, rats, and mice were immunized with either SDS-polyacrylamide gel slices containing denatured fusion proteins or electro-eluted, SDS-free, fusion protein, as well as fusion proteins obtained by other protein purification methods. The presence of antibodies to the fusion proteins in the antisera was assayed by ELISA. Positive serum having a higher titer was further assayed by the double antibody precipitation method using rat ventral prostate cytosol [^{3}H]AR as antigen. The results showed that 1 μl of crude serum precipitated 10 to 20 fmole [^{3}H]AR. Anti-AR crude serum was then affinity-purified by differential suspension of immune serum containing TrpE protein(s) (both those TrpE proteins having and those TrpE proteins not having inserted AR sequences) expressed by pATH vectors. The bound antibodies can be removed from the suspension because TrpE protein is insoluble. Antibodies specific against only the trpE protein were removed; antibodies specific for AR were isolated and again confirmed by both ELISA and double antibody precipitation.

Production of Monoclonal
Anti-Androgen Receptor Antibodies

The immunized rats were judged ready to be sacrificed for a fusion when their serum tested positive anti-AR antibodies by ELISA. Spleens were removed and grinded to release the cells into DMEM (Dulbeco's Modified Engle's Medium) medium. Through a series of centrifugations using DMEM + DMEM with Ficoll Hypaque, the spleen cells were isolated. The SP2/0 myeloma cells were grown, split and diluted in 50 ml of DMEM with 20% FCS, 1% MOPS, and 1X L-Gln for two days before ready for the fusion. SP2/0 cells (5 x 10^6) and 5 x 10^7 spleen cells were used in the fusion. After incubating overnight, the fused cells were collected, suspended in DMEM with 1X H-T, 1X Methotrexate, 20% FCS, and 1X PBS and distributed in 96-well plates. Plates were supplemented after 6 days with DMEM and 20% FCS. Hybridomas were identified and assayed, using the ELISA assay of Engrall, et al., Bio. Chem. et Biophys. ACTA, 251:427-439 (1971). In this assay, plates were coated with either the AR fusion proteins or the TrpE protein as antigen and read on an ELISA reader.

Only those hybridomas that caused a positive reaction with the AR fusion protein were "limit diluted" to a concentration of 10 cells/ml and were then distributed among half of a 96-well plate. The remaining cells from the original well were transferred to a 24-well plate. Each of these plates had a thymocyte feeder layer. The tymocyte feeder layer was made up of thymus cells isolated from an un-injected rat, purified through centrifugation, irradiated with 1200 to 1400 RADS, and diluted to 1 x 10^7 cells/ml of DMEM with 20% FCS.

Positives from these thymocyte 96-well plates were again tested by ELISA. Only those which again

tested positive with the AR fusion protein were grown up for monoclonal antibody purification. Three of the wells produced monoclonal antibody against AR. Both ELISA and double antibody assays were positive. The monoclonal antibodies were designated AN1-6, AN1-7, and AN1-15 and the three cell lines were designated HAN1-6, HAN1-7, and HAN1-15; Accession Nos. 10,000; 9,999; and 10,001; respectively, deposited on January 25, 1989 with the American Type Culture Collection, 12301 Parklawn Drive, Rockville, Maryland 20852.

Specificity of Anti-AR Antibodies

Sucrose gradient centrifugation was used to characterize the specificity of the three monoclonal anti-AR antibodies and their ability to react with non-denatured [^{3}H]AR.

Cytosol was prepared from the ventral prostates of castrated rates as follows. Rats were castrated by the scrotal route while under anesthesia. They were killed 18 hrs. laters by cervical dislocation and their ventral prostates were removed, minced with scissors, washed in Buffer A (50 mM sodium phosphate, pH 7.5, 1 mM EDTA, 2 mM DTT, 10 mM sodium molybdate, 10% (v/v) glycerol and 10 mM sodium floride) and homogenized in 2x the tissue volume of Buffer A + 0.1 mM bacitracin, 1 mM PMSF, and aprotinin (1TIU/ml). The homogenate was centrifuged at 5,000 x g for 10 mins., adjusted to 10 nM ^{3}H-androgen, spun at 225,000 x g for 45 mins. and treated with dextran-coated charcoal. One hundred µl of the cytosol solution, containing 3H-A-AR complexes, was incubated for 6 hrs. with 100 µl of the purified anti-androgen receptor monoclonal antibody, AN1-6, (20x as concentrated as the tissue culture media). Sucrose gradient centrifugation was performed by centrifugation at 257,000 x g for 16 hrs. at 4°C on a 3.8 ml, linear

5-20% (w/v) sucrose gradient containing 20 mM Tris-HCl, pH 7.5, 1 mM EDTA, 1 mM DTT, 10% (v/v) glycerol, and 0.4 M KCl. Gradients were fractionated and numbered from the bottom and 0.2 ml per fraction collected. The results obtained indicated that all three of the monoclonal antibodies, AN1-6, AN1-7, and AN1-15, recognized and effectively bound the radioactively labeled androgen receptor ([^{3}H] AR).

The [^{3}H]AR and other steroid receptor complexes had a sedimendation coefficient of about 4-5S in the sucrose gradient media containing 0.4M KCl. Anti-AR antibodies do not alter the sedimentation coefficient of 4-5S for [^{3}H]glucocorticoid receptors complexes of rat liver, estrogen receptor complexes of MCF-7 cells, and progesterone receptor complexes of T47D cells, but do shift the sedimentation coefficient of [^{3}H]A-AR complexes of rat ventral prostate from 4S to 9-12S or to heavier units. By SDS-polyacrylamide gel electrophoresis analysis it was also found that all major <u>in vitro</u> transcription/translation products of human and rat AR cDNAs were immunoprecipitatable by the anti-AR antibodies.

EXAMPLE 13

Use of AR cDNA and TR2 cDNA as Probes in the Study of Abnormality in Human and Animal Organs and Cancer Cells

Patients with metastatic prostatic cancer initially often respond favorably to androgen withdrawal therapy (castration or antiandrogen treatments). Most patients, however, eventually relapse to an androgen-state for which no chemotherapy, which would significantly increase the survival rate, is available. Regardless of the origin of androgen -independent or -insensitive cancer cells, it is important to understand whether the androgen

insensitivity or abnormality in the diseased cells are due to qualitative or quantitative changes in (a) the AR or TR2 genes, (b) regulation of their transcription, or translation, or (c) other cellular factors. AR cDNA, TR2 cDNA, or their partial segments can be used as specific probes in these studies.

For the analysis of AR or TR2 genes, high molecular weight genomic DNA isolated from target organs, tumors, and cultured cells can be used in identifying and characterizing AR genes. Different restriction endonucleases can be used to cleave DNA. The fragments can be analyzed by Southern analysis (agarose electrophoresis, transfer to nitrocellulose and hybridization with AR cDNA probes). After identification, selected fragments can be cloned and sequenced. It is also possible to use appropriate oligonucleotide fragments of AR or TR2 cDNA as primers to amplify genomic DNA isolated from normal and abnormal organs or cells by specific DNA polymerases. The amplified genomic DNA can then be analyzed to identify sequence abnormality using the polymerase chain reaction (PCR) assay. Saiki, et al., Science, 230, 1350 (1985). See also, Mullis, K.B., U.S. Patent No. 4,683,202; July 28, 1987; and Mullis, K.B., U.S. Patent No. 4,683,195; July 28, 1987. For the analysis of mRNA for ARs or related proteins, dot hybridization and Northern hybridization analysis could be used to characterize mRNA and AR or receptor-like molecules quantitatively and qualitatively. From these studies valuable information about the number of different forms of AR genes and their expression in androgen insensitive and sensitive tumor cells can be obtained.

DNAs and RNAs obtained from androgen sensitive and insensitive tumors and from cell lines from rats and humans with testicular feminization syndromes have been analyzed by the above methods. Preliminary studies

indicated that abnormality in androgen responses may be due to sequence deletion/mutation in genes for ARs.

EXAMPLE 14

Development of Transgenic Animals

Transgenic techniques have been employed for expression of exogenous DNA. It may therefore be possible to confer androgen sensitivity to animals with androgen receptor defects. For example, androgen insensitive animals, such as testicular feminized mice or rats, are known to have defective AR genes or defective AR itself. If DNA containing a normal AR gene is injected into fertilized mouse embryos, the transgenic mice may carry and express the gene and produce a functional AR necessary for androgen responses. For micro-injection, it is necessary to use AR genes containing DNA that can be expressed in the insensitive animals.

A number of genomic receptor clones from human X-chromosome libraries and rat genomic DNA libraries have been obtained and analyzed for their structures. Clones containing AR sequences will be characterized by endonuclease mapping, by Southern hybridization and by S1-nuclease mapping. The 5' and 3' untranslated regions thus identified will aid in determining the minimal size of the DNA that would be required for tissue specific expression of the AR coding region.

Partial sequence analysis of the 5' and 3' regions would locate the minimal region that represents the promoter and the polyadenylation region. Approximately 2 to 5 kb of upstream un-translated region and 0.5 to 1 kb of sequences downstream from the poly(A) site may be fused to the cDNA clone (minimal-gene) and injected into embryos of mice. Transgenic mice would be identified by analysis of their tail DNA using mini-gene specific probe(s).

Normally only some of the transgenic mouse lines can express their transgenes. Transgenes may be inactive because of the presence of inhibitory sequences, integration of the exogenous gene into a transcriptionally inactive chromosomal location, or the juxtaposition of the transgene and an endogenous enhancer. In addition, androgen insensitivity may be due to various other factors and not due to abnormality in the AR gene or its expression.

The foregoing illustrative examples relate to the isolation of human and rat cDNAs encoding DNA binding proteins including androgen receptor and TR-2 and more particularly describe the transcription of the corresponding cDNAs and translation of the corresponding mRNAs in cell-free systems. While the present invention has been described in terms of specific methods and compositions, it is understood that variations and modifications will occur to those skilled in the art upon consideration of the present invention.

Accordingly it is intended in the appended claims to cover all such equivalent variations which come within the scope of the invention as claimed.

WHAT IS CLAIMED IS

1. A purified and isolated DNA sequence encoding androgen receptor polypeptide.

2. The DNA sequence according to claim 1 encoding human androgen receptor polypeptide.

3. The DNA sequence according to claim 1 encoding rat androgen receptor polypeptide.

4. A purified and isolated DNA sequence encoding TR2 polypeptide.

5. The DNA sequence according to claim 1 or 4 which is a cDNA sequence.

6. The DNA sequence according to claim 1 or 4 which is a genomic DNA sequence.

7. The DNA sequence according to claim 1 or 4 which is a partially synthetic DNA sequence.

8. The DNA sequence according to claim 1 and as set forth in Figure 3.

9. The DNA sequence according to claim 4 and as set forth in Figure 4.

10. A procaryotic or eucaryotic host cell transformed or transfected with a DNA sequence according to claim 1 or 4.

11. The procaryotic transformed host cell according to claim 10 which is E. coli DH5α cells designated as, and corresponding to A.T.C.C. deposit Nos.: EC-hAR 3600, A.T.C.C. No. 67879; EC-rAR 2830, A.T.C.C. No. 67878; EC TR2-5, A.T.C.C. 67877; and EC TR2-7, A.T.C.C. No. 67876.

12. A viral or circular DNA plasmid comprising a DNA sequence according to claim 1 or 4.

13. A viral or circular DNA plasmid according to claim 11 further comprising an expression control DNA sequence operatively associated with said androgen receptor or TR2 encoding DNA.

14. A method for the production of androgen receptor polypeptide comprising:

growing, in culture, a host cell transformed or transfected with a DNA sequence according to claim 1; and

isolating from said host cell or culture the polypeptide product of the expression of said DNA sequence.

15. A method for the production of androgen receptor polypeptide comprising:

disposing a DNA sequence according to claim 1 in a cell free transcription and translation system; and

isolating from said system the polypeptide product of the expression of said DNA sequence.

16. A method for the production of TR2 polypeptide comprising:

growing, in culture, a host cell transformed or transfected with a DNA sequence according to claim 4; and

isolating from said host cell or culture the polypeptide product of the expression of said DNA sequence.

17. A method for the production of TR2 polypeptide comprising:

disposing a DNA sequence according to claim 4 in a cell free transcription and translation system; and

isolating from said system the polypeptide product of the expression of said DNA sequence.

18. The polypeptide product of the *in vitro* or *in vivo* expression of a DNA sequence according to claim 1.

19. An amino acid sequence as set out in Figure 3.

20. The polypeptide product of claim 18 characterized by a molecular weights of 98 kD and 79 kD by SDS-PAGE and the ability to bind an androgen.

21. The polypeptide product of the *in vitro* or *in vivo* expression of a DNA sequence according to claim 4.

22. TR2 polypeptides.

23. A synthetic peptide duplicative of a sequence of amino acids present in AR or TR2 proteins in a region of the proteins not involved with DNA binding functions and sharing at least one antigenic epitope with AR or TR2 proteins.

24. An antibody specifically immunoreactive with at least one epitope of androgen receptor polypeptide or TR2 polypeptide other than an epitope within the DNA binding functional region thereof.

25. The monoclonal antibody according to claim 24.

26. The monoclonal antibody according to claim 24 and produced by hybridoma cell line Nos. HB 10,000; HB 9,999; and HB 10,001.

27. The polyclonal antibody according to claim 24.

28. A method for quantitative detection of androgen receptor based on the immunological reaction of androgen receptor with an antibody according to claim 24.

29. A method for quantitative detection of TR2 receptor based on the immunological reaction of TR2 receptor with an antibody according to claim 24.

30. A method for the quantitative detection of androgen receptor encoding DNA or RNA based on hybridization of said nucleic acids with a DNA sequence according to claim 1.

31. A method for the quantitative detection of TR2 receptor encoding DNA or RNA based on hybridization of said nucleic acids with a DNA sequence according to claim 4.

32. A method for the quantitative and qualitative detection of AR or TR2 specific gene sequence or sequences present in a sample comprising the steps of:

a) treating said sample with one oligonucleotide primer for each strand for said specific sequence, under hybridizing conditions such that for each strand of each sequence to which an oligonucleotide primer is hybridized an extension product of each primer is synthesized which is complementary to each nucleic acid strand, wherein said primer or primers are selected so as to be sufficiently complementary to each strand of each specific sequence to hybridize therewith such that the extension product synthesized from one primer, when it is separated from its complement, can serve as a template for synthesis of the extension product of the other primer;

b) treating the sample under denaturing conditions to separate the primer extension products from their templates if the sequence or sequences to be detected are present;

c) treating the sample with oligonucleotide primers such that a primer extension product is synthesized using each of the single strands produced in step (b) as a template, resulting in amplification of the specific nucleic acid sequence or sequences if present;

d) adding to the product of step (c) a labeled oligonucleotide probe for each sequence being detected capable of hybridizing to said sequence or a mutation thereof; and

e) determining whether said hybridization has occurred.

FIGURE 1

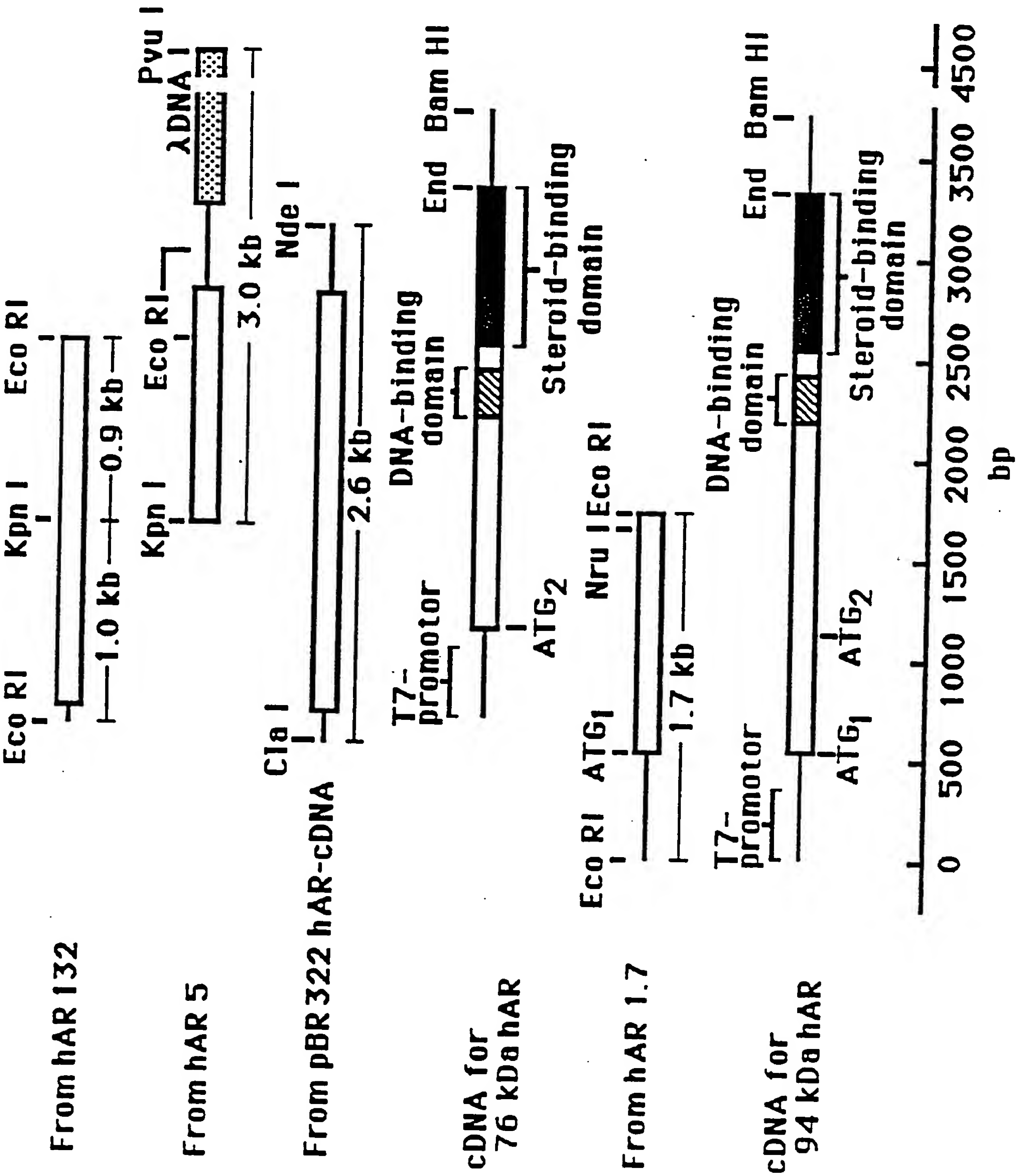
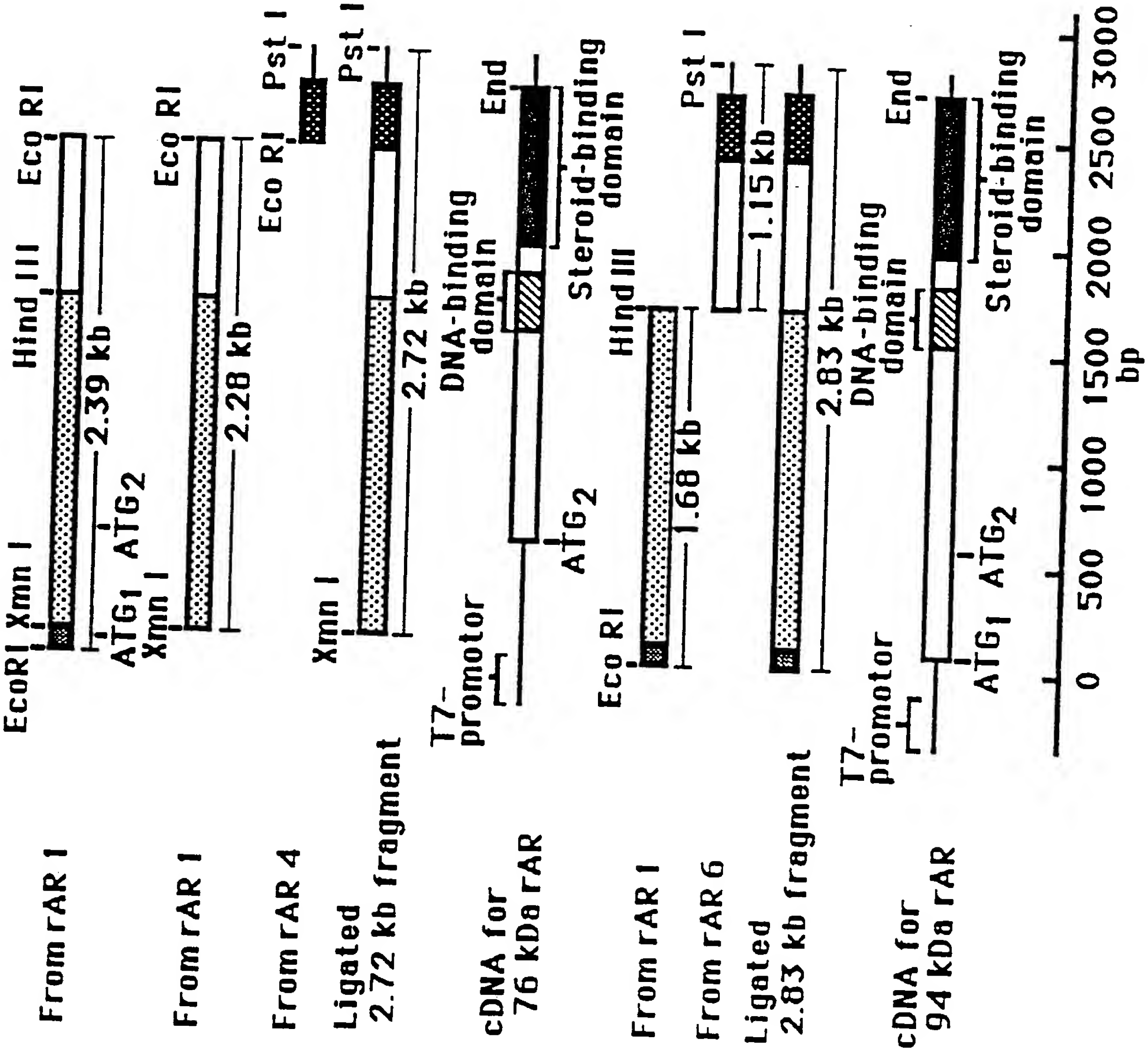

From hAR 132
Eco RI
Kpn I
Eco RI
1.0 kb
0.9 kb
From hAR 5
Kpn I
Eco RI
λDNA
Pvu I
3.0 kb
From pBR 322 hAR-cDNA
Cla I
Nde I
2.6 kb
cDNA for
76 kDa hAR
T7-
promotor
ATG$_2$
DNA-binding
domain
Steroid-binding
domain
End
Bam HI
From hAR 1.7
Eco RI
ATG$_1$
Nru I
Eco RI
1.7 kb
cDNA for
94 kDa hAR
T7-
promotor
ATG$_1$
ATG$_2$
DNA-binding
domain
Steroid-binding
domain
End
Bam HI
0
500
1000
1500
2000
2500
3000
3500
4500
bp

FIGURE 2


From rAR 1
EcoRI
Xmn I
Hind III
Eco RI
2.39 kb
ATG1 ATG2
From rAR 1
Xmn I
Eco RI
2.28 kb
From rAR 4
Eco RI
Pst I
Ligated 2.72 kb fragment
Xmn I
Pst I
2.72 kb
cDNA for 76 kDa rAR
T7-promotor
ATG2
DNA-binding domain
Steroid-binding domain
End
From rAR 1
Eco RI
Hind III
1.68 kb
From rAR 6
Pst I
1.15 kb
Ligated 2.83 kb fragment
2.83 kb
cDNA for 94 kDa rAR
T7-promotor
ATG1 ATG2
DNA-binding domain
Steroid-binding domain
End
0 500 1000 1500 2000 2500 3000
bp

```
                                                              GAATTCCGGCGGAGAGAACCCTCTGTTTTC
TCTTCAGTAGCCAAAAAACAAAACAAACAAAAACAAAAAAGCCGAAATAAAAGAAAAAGATAATAACTCAGTTCTTATTTGCACCTAC
CAAGTATTAAGAGACAGACTGTGAGCCTAGCAGGGCAGATCTTGTCCACCGTGTGTCTTCTTCTGCACGAGACTTTGAGGCTGTCAGA

                                                            GAATTCGGTGGAAGCTAGAGACAAGCTAAAGG
GCATCATCACAGCCTGTTGAACTCTTCTGAGCAAGAGAAGGGGAGGCGGGGTAAGG---G-A--------A-TC--C------C----

Gly Ala Phe Gln Asn Leu Phe Gln Ser Val Arg Glu Ala Ile Gln Asn Pro Gly Pro Arg His Pro
GGA GCG TTC CAG AAT CTG TTC CAG AGC GTG CGC GAA GCG ATC CAG AAC CCG GGC CCC AGG CAC CCT
--- --T --- --- --- --- --- --- --- --- --- --- -T- --- --- --- --- --- --- --- --- --A
 -   -   -   -   -   -   -   -   -   -   -   -  Val  -   -   -   -   -   -   -   -   -

... ... ... ... ... ... ... ... ... ... ... ... Glu Thr Ser Pro Arg Arg Arg Arg Arg Gln
... ... ... ... ... ... ... ... ... ... ... ... GAG ACT AGC CCC CGG CGG CGG CGG CGG CAG
CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAA --- --- --- --- A-- -A- -A- -A- -A- ---
Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln  -   -   -   -   -  Gln Gln Gln Gln  -

Leu Glu Glu Glu Gln Gln Pro Ser Gln Gln Gln Ser Ala Ser Glu Gly His Pro Glu Ser Gly Cys
CTG GAG GAG GAA CAG CAG CCT TCA CAG CAG CAG TCA GCC TCC GAG GGC CAC CCT GAG AGC GGC TGC
--- --T --- --- --- --A --- --- --- -C- --- --G --- CTG --- T-- --- --C --- --A --T ---
 -  Asp  -   -   -   -   -   -   -  Pro  -   -   -  Leu  -  Cys  -   -   -  Arg  -   -
```

FIGURE 3A
TOP RIGHT

```
hAR    1
hAR   31: CCCCACTCTCTCTCCACCTCCTCCTGCCTTCCCCACCCCGAGTGCGGAGCAGAGATCAAAAGATGAAAAGGCAGTCAGG
hAR  198: TTCAGTGGACACTGAATTTGGAAGGTGGAGGATTTTGTTTTTTTCTTTTAAGATCTGGGCATCTTTTGAATCTACCCTT

rAR    1:
hAR  365: GCGCTTTTTGCGTGGTTGCTCCCGCAAGTTTCCTTCTCTGGAGCTTCCCGCAGGTGGGCAGCTAGCTGCAGCGACTACC

rAR    1: Met Glu Val Gln Leu Gly Leu Gly Arg Val Tyr Pro Arg Pro Pro Ser Lys Thr Tyr Arg
rAR   33: ATG GAG GTG CAG TTA GGG CTG GGA AGG GTC TAC CCA CGG CCC CCG TCC AAG ACC TAT CGA
hAR  532: --- --A --- --- --- --- --- --- --- --- --- --T --- --G --- --- --- --- --C ---
hAR    1:  -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -

rAR   43: Glu Ala Ala Ser Ile Ala Pro Pro Gly Ala Cys Leu ... ... ... Gln Gln Arg Gln ...
rAR  159: GAG GCC GCT AGC ATA GCA CCT CCC GGT GCC TGT TTA ... ... ... CAG CAG CGG CAG ...
hAR  658: --- --- --G --- GC- --- --- --- --C --- A-- --G CTG CTG CTG --- --- -A- --- CAG
hAR   43:  -   -   -   -  Ala  -   -   -   -   -  Ser  -  Leu Leu Leu  -   -  Gln  -  Gln

rAR   69: Gln His Pro Glu Asp Gly Ser Pro Gln Ala His Ile Arg Gly Thr Thr Gly Tyr Leu Ala
rAR  237: CAG CAC CCT GAG GAT GGC TCT CCT CAA GCC CAC ATC AGA GGC ACC ACA GGC TAC CTG GCC
hAR  784: --- ... GG- --- --- --T --- --C --- --- --T CGT --- --- C-- --- --- --- --- -T-
hAR   85:  -  ... Gly  -   -   -   -   -   -   -   -  Arg  -   -  Pro  -   -   -   -  Val
```

FIGURE 3A
TOP LEFT

rAR 111:	Leu	Pro	Glu	Pro	Gly	Ala	Ala	Thr	Ala	Pro	Gly	Lys	Gly	Leu	Pro	Gln	Gln	Pro	Pro	Ala
rAR 363:	CTC	CCG	GAG	CCT	GGA	GCT	GCC	ACG	GCT	CCT	GGC	AAG	GGG	CTG	CCG	CAG	CAG	CCA	CCA	GCT
hAR 907:	G--	--A	---	---	---	--C	---	GT-	--C	G-C	A--	---	---	---	---	---	---	-TG	---	--A
hAR 126:	Val	-	-	-	-	-	-	Val	-	Ala	Ser	-	-	-	-	-	-	Leu	-	-
rAR 153:	Leu	Ser	Ser	Cys	Ser	Ala	Asp	Ile	Lys	Asp	Ile	Leu	Ser	Glu	Ala	Gly	Thr	Met	Gln	Leu
rAR 489:	TTA	AGC	AGC	TGC	TCC	GCA	GAC	ATT	AAA	GAC	ATC	CTG	AGC	GAG	GCC	GGC	ACC	ATG	CAA	CTT
hAR1033:	---	---	---	---	---	--T	---	C--	---	---	---	---	---	---	---	A--	---	---	---	--C
hAR 168:	-	-	-	-	-	-	-	Leu	-	-	-	-	-	-	-	Ser	-	-	-	-
rAR 195:	Gln	Glu	Val	Ile	Ser	Glu	Gly	Ser	Ser	Ser	Val	Arg	Ala	Arg	Glu	Ala	Thr	Gly	Ala	Pro
rAR 615:	CAG	GAG	GTA	ATA	TCC	GAA	GGC	AGC	AGC	AGC	GTG	AGA	GCA	AGG	GAG	GCC	ACT	GGG	GCT	CCC
hAR1111:	...	--A	-C-	G--	---	---	---	---	---	---	-G-	---	--G	---	---	---	T-G	---	---	---
hAR 194:	...	-	Ala	Val	-	-	-	-	-	-	Gly	-	-	-	-	-	Ser	-	-	-
rAR 237:	Lys	Ala	Val	Ser	Val	Ser	Met	Gly	Leu	Gly	Val	Glu	Ala	Leu	Glu	His	Leu	Ser	Pro	Gly
rAR 741:	AAA	GCA	GTG	TCT	GTG	TCC	ATG	GGG	TTG	GGT	GTG	GAA	GCA	CTG	GAA	CAT	CTG	AGT	CCA	GGG
hAR1234:	--G	---	---	--G	---	---	---	--C	C-G	---	---	--G	--G	T--	--G	---	---	---	---	---
hAR 235:	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

FIGURE 3A
LEFT BOTTOM

Pro	Pro	Asp	Gln	Asp	Asp	Ser	Ala	Ala	Pro	Ser	Thr	Leu	Ser	Leu	Leu	Gly	Pro	Thr	Phe	Pro	Gly
CCT	CCA	GAT	CAG	GAT	GAC	TCA	GCT	GCC	CCA	TCC	ACG	TTG	TCC	CTA	CTG	GGC	CCC	ACT	TTC	CCA	GGC
---	--G	--C	G--	---	---	---	---	---	---	---	---	---	---	--G	---	---	---	---	---	--C	---
-	-	-	Glu	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Leu	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln
CTT	CAG	CAG	CAG	CAG	CAA	CAG	CAA	CAG	CAG	CAG	CAG	CAG	CAG	CAG	CAG	CAG	CAG	CAG	CAA	CAG	CAG
---	---	--A	---	---	--G	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
-	-	-	-	-	-	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Ser	Ser	Ser	Lys	Asp	Ser	Tyr	Leu	Gly	Gly	Asn	Ser	Thr	Ile	Ser	Asp	Ser	Ala	Lys	Glu	Leu	Cys
TCT	TCC	TCC	AAG	GAT	AGT	TAC	CTA	GGG	GGC	AAT	TCG	ACC	ATA	TCT	GAC	AGT	GCC	AAG	GAG	TTG	TGT
A--	---	---	---	--C	-A-	---	T--	---	---	-C-	---	---	--T	---	---	-AC	---	---	---	---	---
Thr	-	-	-	-	Asn	-	-	-	-	Thr	-	-	-	-	-	Asn	-	-	-	-	-
Glu	Gln	Leu	Arg	Gly	Asp	Cys	Met	Tyr	Ala	Ser	Leu	Leu	Gly	Gly	Pro	Pro	Ala	Val	Arg	Pro	Thr
GAG	CAG	CTT	CGG	GGC	GAC	TGC	ATG	TAC	GCG	TCG	CTC	CTG	GGA	GGT	CCA	CCC	GCC	GTG	CGT	CCC	ACT
--A	---	---	---	--G	--T	---	---	---	--C	C-A	--T	T--	---	-T-	---	---	--T	---	---	---	---
-	-	-	-	-	-	-	-	-	-	Pro	-	-	-	Val	-	-	-	-	-	-	-

FIGURE 3A
RIGHT BOTTOM

```
rAR 279: Pro Cys Ala Pro Leu Ala Glu Cys Lys Gly Leu Ser Leu Asp Glu Gly Pro Gly Lys Gly
rAR 867: CCT TGT GCG CCT CTG GCC GAA TGC AAA GGT CTT TCC CTG GAC GAA GGC CCG GGC AAA GGC
hAR1360: --- --- --C --A T-- --- --- --- --- --- TC- CTG --A --- --C A-- G-A --- --G A--
hAR 277:  -   -   -   -   -   -   -   -   -   -  Ser Leu  -   -  Asp Ser Ala  -   -  Ser

rAR 321: Leu Gly Cys Ser Gly Ser Ser Glu Ala Gly Ser Ser Gly Thr Leu Glu Ile Pro Ser Ser
rAR 993: CTG GGC TGC TCT GGC AGC AGT GAA GCA GGT AGC TCT GGG ACA CTT GAG ATC CCG TCC TCA
hAR1486: --A --- --- --- --- --- GC- -C- --- --G --- --C --- --- --- --A C-G --- --T A-C
hAR 319:  -   -   -   -   -   -  Ala Ala  -   -   -   -   -   -   -   -  Leu  -   -  Thr

rAR 363: Phe Pro Leu Ala Leu Ser Gly Pro Pro His Pro Pro Pro Pro Thr His Pro His Ala Arg
rAR1119: TTT CCG CTC GCT CTG TCC GGG CCG CCG CAC CCC CCG CCC CCT ACC CAT CCA CAC GCC CGC
hAR1612: --- --A --G --- --- G-- --A --- --- -C- --T --- --G --- C-- --- --C --- --T ---
hAR 361:  -   -   -   -   -  Ala  -   -   -  Pro  -   -   -   -  Pro  -   -   -   -   -

rAR 405: Gly Asp Leu Ala Ser Leu His Gly Gly Ser Val Ala Gly Pro Ser Thr Gly Ser Pro Pro
rAR1245: GGG GAC TTG GCT AGC CTA CAT GGA GGG AGT GTA GCC GGA CCC AGC ACT GGA TCG CCC CCA
hAR1738: --- --- C-- --G --- --G --- --C -C- G-- -C- --G --- --- G-T T-- --G --A --- T--
hAR 403:  -   -  Leu  -   -   -   -   -  Ala Gly Ala  -   -   -  Gly Ser  -   -   -  Ser
```

FIGURE 3B
TOP LEFT

```
Thr Glu Glu Thr Ala Glu Tyr Ser Ser Phe Lys Gly Gly Tyr Ala Lys Gly Leu Glu Gly Glu Ser
ACT GAA GAG ACT GCT GAG TAT TCC TCT TTC AAG GGA GGT TAC GCC AAA GGG TTG GAA GGT GAG AGT
--- --- --T --- --- --- --- --- C-- --- --- --- --- --- A-- --- --- C-A --- --C --- --C
 -   -  Asp  -   -   -   -   -  Pro  -   -   -   -   -  Thr  -   -   -   -   -   -   -

Leu Ser Leu Tyr Lys Ser Gly Ala Val Asp Glu Ala Ala Ala Tyr Gln Asn Arg Asp Tyr Tyr Asn
CTG TCT CTG TAT AAG TCT GGA GCA GTA GAC GAG GCA GCA GCA TAC CAG AAT CGC GAC TAC TAC AAC
--- --- --C --C --- --C --- --- C-G --- --- --- --T --G --- --- -G- --- --- --- --- ---
 -   -   -   -   -   -   -   -  Leu  -   -   -   -   -   -   -  Ser  -   -   -   -   -

Ile Lys Leu Glu Asn Pro Leu Asp Tyr Gly Ser Ala Trp Ala Ala Ala Ala Ala Gln Cys Arg Tyr
ATC AAG CTG GAG AAC CCG TTG GAC TAC GGC AGC GCC TGG GCT GCG GCG GCA GCG CAA TGC CGC TAT
--- --- --- --- --- --- C-- --- --- --- --- --- --- --G --T --- --G --- --G --- --- ---
 -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -

Ala Thr Ala Ser Ser Ser Trp His Thr Leu Phe Thr Ala Glu Glu Gly Gln Leu Tyr Gly Pro Gly
GCC ACC GCC TCT TCT TCC TGG CAT ACT CTC TTC ACA GCT GAA GAA GGC CAA TTA TAT GGG CCA GGA
--- G-- --T --C --A --- --- --C --- --- --- --- --C --- --- --- --G --G --- --A --G T-T
 -  Ala  -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -  Cys
```

FIGURE 3B
TOP RIGHT

```
rAR 447:  Gly Gly Gly Gly ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
rAR1371:  GGC GGG GGC GGC ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
hAR1864:  --T --T --T --G GGT GGT GGT GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC
hAR 445:   -   -   -   -  Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly

rAR 471:  Gly Leu Ala Ser Gln Glu Gly Asp Phe Ser Ala Ser Glu Val Trp Tyr Pro Gly Gly Val
rAR1443:  GGG CTG GCA AGC CAG GAG GGT GAC TTC TCT GCC TCT GAA GTG TGG TAT CCT GGT GGA GTT
hAR1990:  --- --- --G G-- --- --A A-C --- --- A-C --A C-- --T --- --- --C --- --C --C A-G
hAR 487:   -   -   -  Gly  -   -  Ser  -   -  Thr  -  Pro Asp  -   -   -   -   -   -  Met

rAR 513:  Tyr Ser Gly Pro Tyr Gly Asp Met Arg Leu Asp Ser Thr Arg Asp His Val Leu Pro Ile
rAR1569:  TAC TCC GGA CCT TAT GGG GAC ATG CGT TTG GAC AGT ACC AGG GAC CAC GTT TTA CCC ATC
hAR2116:  --- --- --- --- --C --- --- --- --- --- --G -C- G-- --- --- --T --- --G --- --T
hAR 529:   -   -   -   -   -   -   -   -   -   -  Glu Thr Ala  -   -   -   -   -   -   -

rAR 555:  Gly Ala Leu Thr Cys Gly Ser Cys Lys Val Phe Phe Lys Arg Ala Ala Glu Gly Lys Gln
rAR1695:  GGA GCT CTC ACT TGT GGC AGC TGC AAG GTC TTC TTC AAA AGA GCT GCG GAA GGG AAA CAG
hAR2242:  --- --- --- --A --- --A --- --- --- --- --- --- --- --- --C --T --- --- --- ---
hAR 571:   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -

rAR 597:  Cys Arg Leu Arg Lys Cys Tyr Glu Ala Gly Met Thr Leu Gly Ala Arg Lys Leu Lys Lys
rAR1821:  TGT CGT CTC CGG AAA TGT TAT GAA GCA GGG ATG ACT CTG GGA GCT CGT AAG CTG AAG AAA
hAR2368:  --- --- --T --- --- --- --- --- --- --- --- --- --- --- --C --G --- --- --- ---
hAR 613:   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -
```

FIGURE 3B
LEFT BOTTOM

```
...  ...  Ser  Ser  Ser  Pro  Ser  Asp  Ala  Gly  Pro  Val  Ala  Pro  Tyr  Gly  Tyr  Thr  Arg  Pro  Pro  Gln
...  ...  AGC  AGT  AGC  CCA  AGC  GAT  GCT  GGG  CCT  GTA  GCC  CCC  TAT  GGC  TAC  ACT  CGG  CCC  CCT  CAG
GGC  GGC  G--  G-C  G--  GGC  G--  --G  --G  -AA  G--  ---  ---  ---  --C  ---  ---  ---  ---  ---  ---  ---
Gly  Gly  Gly  Gly  Gly  Gly  Gly  Glu  -    Glu  Ala  -    -    -    -    -    -    -    -    -    -    -

Val  Asn  Arg  Val  Pro  Tyr  Pro  Ser  Pro  Ser  Cys  Val  Lys  Ser  Glu  Met  Gly  Pro  Trp  Met  Glu  Asn
GTG  AAC  AGA  GTC  CCC  TAT  CCC  AGT  CCC  AGT  TGT  GTT  AAA  AGT  GAA  ATG  GGA  CCT  TGG  ATG  GAG  AAC
---  -G-  ---  --G  ---  ---  ---  ---  ---  -C-  ---  --C  ---  --C  ---  ---  --C  --C  ---  ---  --T  -G-
-    Ser  -    -    -    -    -    -    -    Thr  -    -    -    -    -    -    -    -    -    -    Asp  Ser

Asp  Tyr  Tyr  Phe  Pro  Pro  Gln  Lys  Thr  Cys  Leu  Ile  Cys  Gly  Asp  Glu  Ala  Ser  Gly  Cys  His  Tyr
GAC  TAT  TAC  TTC  CCA  CCC  CAG  AAG  ACC  TGC  CTG  ATC  TGT  GGA  GAT  GAA  GCT  TCT  GGT  TGT  CAC  TAC
---  ---  ---  --T  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  --G  ---  ---  --T
-    -    -    -    -    -    -    -    -    -    -    -    -    -    -    -    -    -    -    -    -    -

Lys  Tyr  Leu  Cys  Ala  Ser  Arg  Asn  Asp  Cys  Thr  Ile  Asp  Lys  Phe  Arg  Arg  Lys  Asn  Cys  Pro  Ser
AAG  TAT  CTA  TGT  GCC  AGC  AGA  AAT  GAT  TGC  ACC  ATT  GAT  AAA  TTT  CGG  AGG  AAA  AAT  TGT  CCA  TCG
---  --C  --G  --C  ---  ---  ---  ---  ---  ---  --T  ---  ---  ---  --C  --A  ---  ---  ---  ---  ---  --T
-    -    -    -    -    -    -    -    -    -    -    -    -    -    -    -    -    -    -    -    -    -

Leu  Gly  Asn  Leu  Lys  Leu  Gln  Glu  Glu  Gly  Glu  Asn  Ser  Ser  Ala  Gly  Ser  Pro  Thr  Glu  Asp  Pro
CTT  GGA  AAT  CTC  AAA  CTA  CAG  GAA  GAA  GGA  GAA  AAC  TCC  AGT  GCT  GGT  AGC  CCC  ACT  GAG  GAC  CCA
---  --T  ---  --G  ---  ---  ---  --G  ---  ---  --G  GCT  ---  --C  A-C  ACC  ---  ---  ---  ---  --G  A--
-    -    -    -    -    -    -    -    -    -    -    -    -    -    Thr  Thr  -    -    -    -    Glu  Thr
```

FIGURE 3B
RIGHT BOTTOM

AR 639:	Ser	Gln	Lys	Met	Thr	Val	Ser	His	Ile	Glu	Gly	Tyr	Glu	Cys	Gln	Pro	Ile	Phe	Leu	Asn	Val	Leu
AR1947:	TCC	CAG	AAG	ATG	ACT	GTA	TCA	CAC	ATT	GAA	GGC	TAT	GAA	TGT	CAA	CCT	ATC	TTT	CTT	AAT	GTC	CTG
AR2494:	A--	---	---	C--	--A	--G	---	---	---	---	---	---	---	---	--G	--C	---	---	--G	---	---	---
AR 655:	Thr	-	-	Leu	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AR 681:	Ala	Ala	Leu	Leu	Ser	Ser	Leu	Asn	Glu	Leu	Gly	Glu	Arg	Gln	Leu	Val	His	Val	Val	Lys	Trp	Ala
AR2073:	GCT	GCC	TTG	TTA	TCT	AGT	CTC	AAC	GAG	CTT	GGC	GAG	AGA	CAG	CTT	GTA	CAT	GTG	GTC	AAG	TGG	GCC
AR2620:	--A	---	---	C-C	---	--C	---	--T	--A	--G	--A	---	---	---	---	---	--C	---	---	---	---	---
AR 697:	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AR 723:	Ser	Trp	Met	Gly	Leu	Met	Val	Phe	Ala	Met	Gly	Trp	Arg	Ser	Phe	Thr	Asn	Val	Asn	Ser	Arg	Met
AR2199:	TCC	TGG	ATG	GGA	CTG	ATG	GTA	TTT	GCC	ATG	GGT	TGG	CGG	TCC	TTC	ACT	AAT	GTC	AAC	TCT	AGG	ATG
AR2746:	---	---	---	--G	--C	---	--G	---	---	---	--C	---	--A	---	---	--C	---	---	---	--C	---	---
AR 739:	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AR 765:	Ser	Gln	Cys	Val	Arg	Met	Arg	His	Leu	Ser	Gln	Glu	Phe	Gly	Trp	Leu	Gln	Ile	Thr	Pro	Gln	Glu
AR2325:	AGC	CAG	TGC	GTG	AGG	ATG	AGG	CAC	CTT	TCT	CAA	GAG	TTT	GGA	TGG	CTC	CAG	ATA	ACC	CCC	CAG	GAA
AR2872:	---	---	--T	--C	C-A	---	---	---	--C	---	---	---	---	---	---	---	--A	--C	---	---	---	---
AR 781:	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

FIGURE 3C
TOP LEFT

```
Glu Ala Ile Leu Glu Ala Ile Glu Pro Gly Val Val Cys Ala Gly His Asp Asn Asn Gln Pro Asp Ser Phe
GAA GCC ATT CTG GAA GCC ATT GAG CCA GGA GTG GTG TGT GCC GGA CAT GAC AAC AAC CAG CCT GAT TCC TTT
--- --- --- --- --- --- --- --- --- --T --A --- --- --T --- --C --- --- --- --- --C --C --- ---
 -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -

Lys Ala Leu Ala Lys Ala Leu Pro Gly Phe Arg Asn Leu His Val Asp Asp Gln Met Ala Val Ile Gln Tyr
AAG GCC TTG GCC AAG GCC TTG CCT GGC TTC CGC AAC TTG CAT GTG GAT GAC CAG ATG GCA GTC ATT CAG TAT
--- --- --- --- --- --- --- --- --- --- --- --- --A --C --- --C --- --- --- --T --- --- --- --C
 -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -

Leu Tyr Phe Met Leu Tyr Phe Ala Pro Asp Leu Val Phe Asn Glu Tyr Arg Met His Lys Ser Arg Met Tyr
CTC TAC TTT ATG CTC TAC TTT GCA CCT GAC CTG GTT TTC AAT GAG TAT CGC ATG CAC AAG TCT CGA ATG TAC
--- --- --C --- --- --- --C --C --- --T --- --- --- --- --- --C --- --- --- --- --C --G --- ---
 -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -

Phe Leu Cys Glu Phe Leu Cys Met Lys Ala Leu Leu Leu Phe Ser Ile Ile Pro Val Asp Gly Leu Lys Asn
TTC CTG TGC GAA TTC CTG TGC ATG AAA GCA CTG CTA CTC TTC AGC ATT ATT CCA GTG GAT GGG CTG AAA AAT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -
```

FIGURE 3C
TOP RIGHT

```
rAR 807: Gln Lys Phe Phe Asp Glu Leu Arg Met Asn Tyr Ile Lys Glu Leu Asp Arg Ile Ile Ala
rAR2451: CAA AAA TTC TTT GAT GAA CTT CGA ATG AAC TAC ATC AAG GAA CTT GAT CGC ATC ATT GCA
hAR2998: --- --- --- --- --- --- --- --- --- --- --- --- --- --- --C --- --T --- --- ---
hAR 823:  -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -

rAR 849: Val Gln Pro Ile Ala Arg Glu Leu His Gln Phe Thr Phe Asp Leu Leu Ile Lys Ser His
rAR2577: GTG CAG CCT ATT GCA AGA GAG CTG CAT CAA TTC ACT TTT GAC CTG CTA ATC AAG TCC CAT
hAR3124: --- --- --- --- --G --- --- --- --- --G --- --- --- --- --- --- --- --- --A --C
hAR 865:  -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -

rAR 891: Ser Gly Lys Val Lys Pro Ile Tyr Phe His Thr Gln
rAR2703: TCT GGG AAA GTC AAG CCC ATC TAT TTC CAC ACA CAG TGA AGATTTGGAAACCCTAATACCCAAACC
hAR3250: --- --- --- --- --- --- --- --- --- --- --C --- --- --CA-----------T-T---C-C--
hAR 907:  -   -   -   -   -   -   -   -   -   -   -   -

rAR2857: TGTACAGTCTGTCATGAACATGTTCCCCAGTTCTATTTCCTGGGCTTTTCCTTCTTTCTTTTTCTTCTTCTCTGCCTCT
hAR3404: -AC-GTC-G-CATGGA-TTC-A--TG-TG-GCT-T---TT-CTCT--C--TCCT--CT----CT-CT-C-CTCC-TATC

rAR3024: TGATCTTCTTGTGGCCCAGTGTCAACTGTGCTTGTTTATAGCACTGTGCTGTGTGCCAACCAAGCAAATGTTTACTCAC
hAR3571: -----C--A-A--------------G-TGTGC-TG--TACAGCACTACTCTGTGC-AGC-AC-CA--CGT--ACT-AT-

rAR3191: ACCAAAAAGCAAAAAAAAAAGGAATTC
```

FIGURE 3C
LEFT BOTTOM

Cys	Lys	Arg	Lys	Asn	Pro	Thr	Ser	Cys	Ser	Arg	Arg	Phe	Tyr	Gln	Leu	Thr	Lys	Leu	Leu	Asp	Ser
TGC	AAA	AGA	AAA	AAT	CCC	ACA	TCC	TGC	TCA	AGG	CGC	TTC	TAC	CAG	CTC	ACC	AAG	CTC	CTG	GAT	TCT
---	---	---	---	---	---	---	---	---	---	--A	---	---	---	---	---	---	---	---	---	--C	--C
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Met	Val	Ser	Val	Asp	Phe	Pro	Glu	Met	Met	Ala	Glu	Ile	Ile	Ser	Val	Gln	Val	Pro	Lys	Ile	Leu
ATG	GTG	AGC	GTG	GAC	TTT	CCT	GAA	ATG	ATG	GCA	GAG	ATC	ATC	TCT	GTG	CAA	GTG	CCC	AAG	ATC	CTT
---	---	---	---	---	---	--G	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

CACCTTGTTCCCTTTTCAGATGTCTTCTGCCTGTTATATAACTCTGCACTACTTCTCTGCAGTGCCTTGGGGAAATTCCTCTACTGA
--G--CA-G---CC--TCAGATGTC-TCTG-CTG-TATA-CTCTGCACTACTCCTCTGCAGTGC-T-G---AATTTCCT--A-TGATG

TTTACCCTCCCATGGCACATTTTGAATCCGCTGCGTGTTGTGCTCCTGCCTGTGTTTTGAGTTTTGTTGTATTTCTTCAAGTCTGTGA
-AAC--TC--ATG-CAC-T-CAGACT-TGCT-C-CAT-GTG------AT-----------A-GG--------GC---T--A-------

CTTATGCCATGGCAAGTTTAGAGAGCTATAAGTATCTTGGGAAGAAACAAACAGAGAGAGTAAAAAAACCAAAAAACAAAACAAAAAA
T-ATGC-ACG--A-GT--AGAGAGCTA-G-TTATCTGG--A--TC--A-C-A-A-C-CCCG--TTC

FIGURE 3C
RIGHT BOTTOM

```
     GGCCCGTCGGCTTTCTTCAACCCTCTCTTCCCGGAGCGCCCCCAATCCACGAGTGGCAGCCGCG

     ATG GCA ACC ATA GAA GAA ATT GCA CAT CAA ATT ATT GAA CAA CAG ATG
  1: Met Ala Thr Ile Glu Glu Ile Ala His Gln Ile Ile Glu Gln Gln Met

     ATT GTG ACA GCA CTT GAT CAT AAT ACC CAA GGC AAG CAG TTC ATT CTG
 31: Ile Val Thr Ala Leu Asp His Asn Thr Gln Gly Lys Gln Phe Ile Leu

     AGG CAA GAT TCC ACT CCG GGA AAA GTT TTC CTT ACA ACT CCA GAT GCA
 61: Arg Gln Asp Ser Thr Pro Gly Lys Val Phe Leu Thr Thr Pro Asp Ala

     GCA CAA CAC CTG CAG CTC CTA ACA GAT AAT TCT CCA GAC CAA GGA CCA
 91: Ala Gln His Leu Gln Leu Leu Thr Asp Asn Ser Pro Asp Gln Gly Pro

     TCA GGA CGT CAT TAT GGA GCA GTA ACT TGT GAA GGC TGC AAA GGA TTT
121: Ser Gly Arg His Tyr Gly Ala Val Thr Cys Glu Gly Cys Lys Gly Phe

     GGA TCA AAG GAT TGT ATT ATT AAT AAG CAC CAC CGA AAC CGC TGT CAA
151: Gly Ser Lys Asp Cys Ile Ile Asn Lys His His Arg Asn Arg Cys Gln

     GAC TCT GTC CAA TGT GAA AGA AAA CCC ATT GAA GTA TCA CGA GAA AAA
181: Asp Ser Val Gln Cys Glu Arg Lys Pro Ile Glu Val Ser Arg Glu Lys

     AAG GAC CTT CGT AGC CCA TTA ACT GCA ACT CCA ACT TTT GTA ACA GAT
211: Lys Asp Leu Arg Ser Pro Leu Thr Ala Thr Pro Thr Phe Val Thr Asp

     TTC ATG AAT ATT CAT CCA TCT GGA GTA AAA ACT GAG TCA GCT GTG CTG
241: Phe Met Asn Ile His Pro Ser Gly Val Lys Thr Glu Ser Ala Val Leu

     ACA TTG GCC AAT GTG GTT ACA TCA TTA GCG AAT CTT GGA AAA ACT AAA
271: Thr Leu Ala Asn Val Val Thr Ser Leu Ala Asn Leu Gly Lys Thr Lys

     TTA AGC AAT GAT GAT ACC TCT TTG TGT GAA TTT CAA GAA ATG CAG ACC
301: Leu Ser Asn Asp Asp Thr Ser Leu Cys Glu Phe Gln Glu Met Gln Thr
```

FIGURE 4A
LEFT

GAATTCG 7

GGACTGTCGCGTCGGCGCCCGACGCGGAGTCAGCAGGGGCGAAAAGCGGTAGATC 126

GGA GAG ATT GTT ACA GAG CAG CAA ACT GGG CAG AAA ATC CAG 216

Gly Glu Ile Val Thr Glu Gln Gln Thr Gly Gln Lys Ile Gln

ACA AAT CAC GAC GGC TCT ACT CCA AGC AAA GTC ATT CTG GCC 306

Thr Asn His Asp Gly Ser Thr Pro Ser Lys Val Ile Leu Ala

GCA GGT GTC AAC CAG TTA TTT TTT ACC ACT CCT GAT CTG TCT 396

Ala Gly Val Asn Gln Leu Phe Phe Thr Thr Pro Asp Leu Ser

AAT AAG GTT TTT GAT CTT TGC GTA GTA TGT GGA GAC AAA GCA 486

Asn Lys Val Phe Asp Leu Cys Val Val Cys Gly Asp Lys Ala

TTT AAA AGA AGC ATC CGA AAA AAT TTA GTA TAT TCA TGT CGA 576

Phe Lys Arg Ser Ile Arg Lys Asn Leu Val Tyr Ser Cys Arg

TAC TGC AGG TTA CAG AGA TGT ATT GCG TTT GGA ATG AAG CAA 666

Tyr Cys Arg Leu Gln Arg Cys Ile Ala Phe Gly Met Lys Gln

TCT TCC AAC TGT GCC GCT TCA ACA GAA AAA ATC TAT ATC CGA 756

Ser Ser Asn Cys Ala Ala Ser Thr Glu Lys Ile Tyr Ile Arg

AGT GAA AGT ACA AGG TCA ACA GGA CTG TTA GAT TCA GGA ATG 846

Ser Glu Ser Thr Arg Ser Thr Gly Leu Leu Asp Ser Gly Met

ATG ACA TCA GAT AAG GCT GAA TCA TGT CAG GGA GAT TTA AGT 936

Met Thr Ser Asp Lys Ala Glu Ser Cys Gln Gly Asp Leu Ser

GAT CTT TCT CAA AAT AGT AAT GAA ATG TCT ATG ATT GAA AGC 1026

Asp Leu Ser Gln Asn Ser Asn Glu Met Ser Met Ile Glu Ser

AAC GGT GAT GTT TCA AGG GCA TTT GAC ACT CTT GCA AAA GCA 1116

Asn Gly Asp Val Ser Arg Ala Phe Asp Thr Leu Ala Lys Ala

FIGURE 4A
RIGHT

```
      TTG AAT CCT GGA GAG AGC ACA GCC TGC CAG AGC TCA GTA GCG GGC
331:  Leu Asn Pro Gly Glu Ser Thr Ala Cys Gln Ser Ser Val Ala Gly

      TAC ACC GAA AAA GAG GGG CCA CTT CTC AGC GAT TCA CAT GTA GCT
361:  Tyr Thr Glu Lys Glu Gly Pro Leu Leu Ser Asp Ser His Val Ala

      CAC TAC ATT GGG GAG TCT GCC TCC AGA CTG CTG TTC TTA TCA ATG
391:  His Tyr Ile Gly Glu Ser Ala Ser Arg Leu Leu Phe Leu Ser Met

      AAC AGC ATA TCA CTG GTG AAA GCT TAC TGG AAT GAA CTT TTT ACT
421:  Asn Ser Ile Ser Leu Val Lys Ala Tyr Trp Asn Glu Leu Phe Thr

      TTA GCA ACA TTT GTC AAT TGT CTT CAC AAT AGT CTT CAA CAA GAT
451:  Leu Ala Thr Phe Val Asn Cys Leu His Asn Ser Leu Gln Gln Asp

      ACT GAT TTA TAA ATGCTTAACTATAGAATGGCTTATGACTACCCAAAACAGTGCCC
481:  Thr Asp Leu ***

      ACCTTTTAGTTCTTTAGCATATTCTATTTCTTATTGTTTTATATAATTTTTAAATCATTT

      ATCTACTAGAAGGCAGCATCACATTCCCATCTTACTTATGGACTCCTACCCCTGGTTCAT

      TACTAATGGATGGGTTTTAAACATGTCCCTCTACAATAAATTAAAATCTTTCAATGTTTG
```

*: 11 of 30 TR2 clones have extra 429 bp insert here which

```
GTATGTATTAGCTTTTAAGGAGAAAATACTTTTTAAAGATTCCAGCAAACTACAAGAGT
AAATATAAGAATACAGACCTACGTATGTACTTTTTTATTTATTCAACTAATATTTATTG
AAATTGGTTTGTTTTACTAATCTAAAGCAACTTTGTTGAACTTGCACATAATTTCTAAA
CTTACCTACAGCTTACATTTCTAATTGTCTGTAATCCTATATTGTGATATAATAGTTTA
```

FIGURE 4B
LEFT

ATG GAA GGA AGT GTA CAC CTA ATC ACT GGA GAT TCA AGC ATA AAT 1206
Met Glu Gly Ser Val His Leu Ile Thr Gly Asp Ser Ser Ile Asn

TTC AGG CTC ACC ATG CCT TCT CCT ATG CCT GAG TAC CTG AAT GTG 1296
Phe Arg Leu Thr Met Pro Ser Pro Met Pro Glu Tyr Leu Asn Val

CAC TGG GCA CTT TCG ATT CCT TCT TTC CAG GCT CTA GGG CAA GAA 1386
His Trp Ala Leu Ser Ile Pro Ser Phe Gln Ala Leu Gly Gln Glu

CTT GGT CTT GCC CAG TGC TGG CAA GTG ATG AAT GTA GCA ACT ATA 1476
Leu Gly Leu Ala Gln Cys Trp Gln Val Met Asn Val Ala Thr Ile

GCC AAG GTA ATT GCA GCC CTC ATT CAT TTC ACA AGA CGA GCA ATC 1566
Ala Lys Val Ile Ala Ala Leu Ile His Phe Thr Arg Arg Ala Ile

CATCAACAAATGGGGAAAATTGCCTTTTGAGCTCAGGAATAATTTATAAATTGGGGACT 1681

GCTTCCTCCTTATGTTTAACAGCAGAGGGGTAATCACCTTAAAATGTCATCAAAAATAG 1800

GTCTTATATGCCTGTAATGGTTATAAAGCCTACCTTCAGGAAAGCTATGGTTGACTAAT 1919

AATATAATGTGGAGGTGTTTACCTGAGGGCCTCTCTATCTCCCCGAATTC 2029

create a termination codon TAG.
ATTGAAATTAACAAAATATGTCAAATATGTATACTTTTTAGTTTACAGTTTTCCAACTA
AATACCTGTGTAAAAAGCACGTATTTGTAGTTTGTGTTTGGCAAAAAGAATGAGAGATGC
GATTGATGGTTATCTTTGGAGTTTAGTATGGTAGCCATGTCTCCTATTAGCAGCATTAAG
ACACATTTTTGTAG

FIGURE 4B
RIGHT

h-GR 419	Lys	Leu	Cys	Leu	Val	Cys	Ser	Asp	Glu	Ala	Ser
h-MR 601	Lys	Ile	Cys	Leu	Val	Cys	Gly	Asp	Glu	Ala	Ser
h-PR 565	Lys	Ile	Cys	Leu	Ile	Cys	Gly	Asp	Glu	Ala	Ser
h-AR	Lys	Thr	Cys	Leu	Ile	Cys	Gly	Asp	Glu	Ala	Ser
r-AR	Lys	Thr	Cys	Leu	Ile	Cys	Gly	Asp	Glu	Ala	Ser
h-ER 183	Arg	Tyr	Cys	Ala	Val	Cys	Asn	Asp	Tyr	Ala	Ser
h-TR2	Asp	Leu	Cys	Val	Val	Cys	Gly	Asp	Lys	Ala	Ser
v-erbA35	Glu	Gln	Cys	Val	Val	Cys	Gly	Asp	Lys	Ala	Thr
c-VDR	Arg	Ile	Cys	Gly	Val	Cys	Gly	Asp	Arg	Ala	Thr

FIGURE 5
TOP LEFT

Gly	Cys	His	Tyr	Gly	Val	Leu	Thr	Cys	Gly	Ser	Cys	Lys
Gly	Cys	His	Tyr	Gly	Val	Val	Thr	Cys	Gly	Ser	Cys	Lys
Gly	Cys	His	Tyr	Gly	Val	Leu	Thr	Cys	Gly	Ser	Cys	Lys
Gly	Cys	His	Tyr	Gly	Ala	Leu	Thr	Cys	Gly	Ser	Cys	Lys
Gly	Cys	His	Tyr	Gly	Ala	Leu	Thr	Cys	Gly	Ser	Cys	Lys
Gly	Tyr	His	Tyr	Gly	Val	Trp	Ser	Cys	Glu	Gly	Cys	Lys
Gly	Arg	His	Tyr	Gly	Ala	Val	Thr	Cys	Glu	Gly	Cys	Lys
Gly	Tyr	His	Tyr	Arg	Cys	Ile	Thr	Cys	Glu	Gly	Cys	Lys
Gly	Phe	His	Phe	Asn	Ala	Met	Thr	Cys	Glu	Gly	Cys	Lys

FIGURE 5
TOP RIGHT

h-GR 439	Val	Phe	Phe	Lys	Arg	Ala	Val	Glu	Gly	Gln	His
h-MR 625	Val	Phe	Phe	Lys	Arg	Ala	Val	Glu	Gly	Gln	His
h-PR 589	Val	Phe	Phe	Lys	Arg	Ala	Met	Glu	Gly	Gln	His
h-AR	Val	Phe	Phe	Lys	Arg	Ala	Ala	Glu	Gly	Lys	Gln
r-AR	Val	Phe	Phe	Lys	Arg	Ala	Ala	Glu	Gly	Lys	Gln
h-ER 207	Ala	Phe	Phe	Lys	Arg	Ser	Ile	Gln	Gly	His	Asn
h-TR2	Gly	Phe	Phe	Lys	Arg	Ser	Ile	Arg	Lys	Asn	Leu
v-erbA59	Ser	Phe	Phe	Arg	Arg	Thr	Ile	Gln	Lys*	His	Pro
c-VDR	Gly	Phe	Phe	Arg	Arg	Ser	Met	Lys	Arg	Lys	Ala

FIGURE 5
MIDDLE LEFT

Asn	Tyr	Leu	Cys	Ala	Gly	Arg	Asn	Asp	Cys	Ile	Ile	Asp
Asn	Tyr	Leu	Cys	Ala	Gly	Arg	Asn	Asp	Cys	Ile	Ile	Asp
Asn	Tyr	Leu	Cys	Ala	Gly	Arg	Asn	Asp	Cys	Ile	Val	Asp
Lys	Tyr	Leu	Cys	Ala	Ser	Arg	Asn	Asp	Cys	Thr	Ile	Asp
Lys	Tyr	Leu	Cys	Ala	Ser	Arg	Asn	Asp	Cys	Thr	Ile	Asp
Asp	Tyr	Met	Cys	Pro	Ala	Thr	Asn	Gln	Cys	Thr	Ile	Asp
Val	Tyr	Ser	Cys	Arg	Gly	Ser	Lys	Asp	Cys	Ile	Ile	Asn
Thr	Tyr	Ser	Cys	Thr	Tyr	Asp	Gly	Cys	Cys	Val	Ile	Asp
Met	Phe	Thr	Cys	Pro	Phe	Asn	Gly	Asp	Cys	Lys	Ile	Thr

FIGURE 5
MIDDLE RIGHT

h-GR 463	Lys	Ile	Arg	Arg	Lys	Asn	Cys	Pro	Ala	Cys	Arg
h-MR 649	Lys	Ile	Arg	Arg	Lys	Asn	Cys	Pro	Ala	Cys	Arg
h-PR 613	Lys	Ile	Arg	Arg	Lys	Asn	Cys	Pro	Ala	Cys	Arg
h-AR	Lys	Phe	Arg	Arg	Lys	Asn	Cys	Pro	Ser	Cys	Arg
r-AR	Lys	Phe	Arg	Arg	Lys	Asn	Cys	Pro	Ser	Cys	Arg
h-ER 231	Lys	Asn	Arg	Arg	Lys	Ser	Cys	Gln	Ala	Cys	Arg
h-TR2	Lys	His	His	Arg	Asn	Arg	Cys	Gln	Tyr	Cys	Arg
v-erbA85	Lys	Ile	Thr	Arg	Asn	Gln	Cys	Gln	Leu	Cys	Arg
c-VDR	Lys	Asp	Asn	Arg	Arg	His	Cys	Gln	Ala	Cys	Arg

FIGURE 5
BOTTOM LEFT

Tyr	Arg	Lys	Cys	Leu	Gln	Ala	Gly	Met	Asn	Leu	Glu	Ala		
Leu	Gln	Lys	Cys	Leu	Gln	Ala	Gly	Met	Asn	Leu	Gly	Ala		
Leu	Arg	Lys	Cys	Cys	Gln	Ala	Gly	Met	Val	Leu	Gly	Gly		
Leu	Arg	Lys	Cys	Tyr	Glu	Ala	Gly	Met	Thr	Leu	Gly	Ala		
Leu	Arg	Lys	Cys	Tyr	Glu	Ala	Gly	Met	Thr	Leu	Gly	Ala		
Leu	Arg	Lys	Cys	Tyr	Glu	Val	Gly	Met	Met	Lys	Gly	Gly		
Leu	Gln	Arg	Cys	Ile	Ala	Phe	Gly	Met	Lys	Gln	Asp	Cys		
Phe	Lys	Lys	Cys	Ile	Ser	Val	Gly	Met	Ala	Met	Asp	Leu		
Leu	Lys	Arg	Cys	Val	Asp	Ile	Gly	Met	Met	Lys				

FIGURE 5
BOTTOM RIGHT

```
         170                 180                 190                 200
ATG CAA ACA CAA AAA CCG ACT CTC GAA CTG CTA ACC TGC GAA GGC
MET GLN THR GLN LYS PRO THR LEU GLU LEU LEU THR CYS GLU GLY
  1                                      10
         260                 270                 280                 290
GAT CGT CCG GCA ACG CTG CTG CTG GAA TCC GCA GAT ATC GAC AGC
ASP ARG PRO ALA THR LEU LEU LEU GLU SER ALA ASP ILE ASP SER
                                         40
         350                 360                 370                 380
ATT ACA GCT TTA GGT GAC ACT GTC ACA ATC CAG GCA CTT TCC GGC
ILE THR ALA LEU GLY ASP THR VAL THR ILE GLN ALA LEU SER GLY
                                         70
         440                 450                 460                 470
GGT GTG GAA AGT GAA CAA TCA CCA AAC TGC CGT GTG CTG CGC TTC
GLY VAL GLU SER GLU GLN SER PRO ASN CYS ARG VAL LEU ARG PHE
                                        100
         530                 540                 550                 560
CTT TCG GTT TTT GAC GCT TTC CGT TTA TTG CAG AAT CTG TTG AAT
LEU SER VAL PHE ASP ALA PHE ARG LEU LEU GLN ASN LEU LEU ASN
                                        130
```

FIGURE 6A
TOP LEFT

323 amino acids from TRP E protein

```
210                 220                 230                 240                 250
GCT TAT CGC GAC AAT CCC ACC GCG CTT TTT CAC CAG TTG TGT GCG
ALA TYR ARG ASP ASN PRO THR ALA LEU PHE HIS GLN LEU CYS GLY
                     20                                      30
300                 310                 320                 330                 340
AAA GAT GAT TTA AAA AGC CTG CTG CTG GTA GAC AGT GCG CTG CGC
LYS ASP ASP LEU LYS SER LEU LEU LEU VAL ASP SER ALA LEU ARG
                     50                                      60
390                 400                 410                 420                 430
AAC GGC GAA GCC CTC CTG GCA CTA CTG GAT AAC GCC CTG CCT GCG
ASN GLY GLU ALA LEU LEU ALA LEU LEU ASP ASN ALA LEU PRO ALA
                     80                                      90
480                 490                 500                 510                 520
CCC CCT GTC AGT CCA CTG CTG GAT GAA GAC GCC CGC TTA TGC TCC
PRO PRO VAL SER PRO LEU LEU ASP GLU ASP ALA ARG LEU CYS SER
                    110                                     120
570                 580                 590                 600                 610
GTA CCG AAG GAA GAA CGA GAA GCC ATG TTC TTC AGC GGC CTG TTC
VAL PRO LYS GLU GLU ARG GLU ALA MET PHE PHE SER GLY LEU PHE
                    140                                     150
```

FIGURE 6A
TOP RIGHT

```
          620                 630                  640                 650
TCT TAT GAC CTT GTG GCG GGA TTT GAA GAT TTA CCG CAA CTG TCA
SER TYR ASP LEU VAL ALA GLY PHE GLU ASP LEU PRO GLN LEU SER
                                        160
          710                 720                  730                 740
CTG ATG GTG ATT GAC CAT CAG AAA AAA AGC ACC CGT ATT CAG GCC
LEU MET VAL ILE ASP HIS GLN LYS LYS SER THR ARG ILE GLN ALA
                                        190
          800                 810                  820                 830
CGC CTG AAC GAA CTA CGT CAG CAA CTG ACC GAA GCC GCG CCG CCG
ARG LEU ASN GLU LEU ARG GLN GLN LEU THR GLU ALA ALA PRO PRO
                                        220
          890                 900                  910                 920
AGC GAT GAA GAG TTC GGT GGC GTA GTG CGT TTG TTG CAA AAA GCG
SER ASP GLU GLU PHE GLY GLY VAL VAL ARG LEU LEU GLN LYS ALA
                                        250
          980                 990                 1000                1010
TCT CTG CCC TGC CCG TCA CCG CTG GCG GCC TAT TAC GTG CTG AAA
SER LEU PRO CYS PRO SER PRO LEU ALA ALA TYR TYR VAL LEU LYS
                                        280
         1070                1080                 1090                1100
TTC ACC CTA TTT GGC GCG TCG CCG GAA AGC TCG CTC AAG TAT GAT
PHE THR LEU PHE GLY ALA SER PRO GLU SER SER LEU LYS TYR ASP
                                        310
```

FIGURE 6A
BOTTOM LEFT

```
660                 670                 680                 690                 700
GCG GAA AAT AAC TGC CCT GAT TTC TGT TTT TAT CTC GCT GAA ACG
ALA GLU ASN ASN CYS PRO ASP PHE CYS PHE TYR LEU ALA GLU THR
                170                                                             180
750                 760                 770                 780                 790
AGC CTG TTT GCT CCG AAT GAA GAA GAA AAA CAA CGT CTC ACT GCT
SER LEU PHE ALA PRO ASN GLU GLU GLU LYS GLN ARG LEU THR ALA
                200                                                             210
840                 850                 860                 870                 880
CTG CCA GTG GTT TCC GTG CCG CAT ATG CGT TGT GAA TGT AAT CAG
LEU PRO VAL VAL SER VAL PRO HIS MET ARG CYS GLU CYS ASN GLN
                230                                                             240
930                 940                 950                 960                 970
ATT CGC GCT GGA GAA ATT TTC CAG GTG GTG CCA TCT CGC CGT TTC
ILE ARG ALA GLY GLU ILE PHE GLN VAL VAL PRO SER ARG ARG PHE
                260                                                             270
1020                1030                1040                1050                1060
AAG AGT AAT CCC AGC CCG TAC ATG TTT TTT ATG CAG GAT AAT GAT
LYS SER ASN PRO SER PRO TYR MET PHE PHE MET GLN ASP ASN ASP
                290                                                             300
1110                1120                1130
GCC ACC AGC CGC CAG ATT GAG ATC
ALA THR SER ARG GLN ILE GLU ILE
                320
```

FIGURE 6A
BOTTOM RIGHT

6 amino acid linker

```
CCC CGG AAT TCG AGC TCG
Pro Arg Asn Ser Ser Ser
```

GGGACACTTGAACTGCCGTCTACCCTGTCTCTC
GlyThrLeuGluLeuProSerThrLeuSerLeu

CCACTGGCTCTGGCCGGACCGCCGCCCCTCCGCCGCCTCCCCATCCCCACGCTCGCATC
ProLeuAlaLeuAlaGlyProProProProProProProProHisProHisAlaArgIle

TATGGGGACCTGGCGAGCCTGCATGGCGCGGGTGCAGCGGGACCCGGTTCTGGGTCACCC
TyrGlyAspLeuAlaSerLeuHisGlyAlaGlyAlaAlaGlyProGlySerGlySerPro

GGACCGTGTGGTGGTGGTGGGGGTGGTGGTGGCGGCGGCGGCGGCGGCGGCGGCGGC
GlyProCysGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly

ACTCGGCCCCCTCAGGGGCTGGCGGGCCAGGAAAGCGACTTCACCGCACCTGATGTGTGG
ThrArgProProGlnGlyLeuAlaGlyGlnGluSerAspPheThrAlaProAspValTrp

ATGGGCCCCTGGATGGATAGCTACTCCGGACCTTACGGGGACATGCGTTTGGAGACTGCC
MetGlyProTrpMetAspSerTyrSerGlyProTyrGlyAspMetArgLeuGluThrAla

GGAGATGAAGCTTCTGGGTGTCACTATGGAGCT
GlyAspGluAlaSerGlyCysHisTyrGlyAla

CGC CCG GGG ATC CTC TAG
Arg Pro Gly Ile Leu STOP

Total amino acid: 323 + 6 + 242 + 5 = 576

FIGURE 6B
LEFT

TACAAGTCCGGAGCACTGGACGAGGCAGCTGCGTACCAGAGTCGCGACTACTACAACTTT
TyrLysSerGlyAlaLeuAspGluAlaAlaAlaTyrGlnSerArgAspTyrTyrAsnPhe

AAGCTGGAGAACCCGCTGGACTACGGCAGCGCCTGGGCGGCTGCGGCGCAGTGCCGC
LysLeuGluAsnProLeuAspTyrGlySerAlaTrpAlaAlaAlaAlaGlnCysArg

TCAGCCGCCGCTTCCTCATCCTGGCACACTCTCTTCACAGCCGAAGAAGGCCAGTTGTAT
SerAlaAlaAlaSerSerSerTrpHisThrLeuPheThrAlaGluGluGlyGlnLeuTyr

GGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGAGGCGGAAGCTGTAGCCCCCTACGGCTAC
GlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGluAlaGluAlaValAlaProTyrGlyTyr

TACCCTGGCGGCATGGTGAGCAGAGTGCCCTATCCCAGTCCCACTTGTGTCAAAAGCGAA
TyrProGlyGlyMetValSerArgValProTyrProSerProThrCysValLysSerGlu

AGGGACCATGTTTTGCCCATTGACTATTACTTCCACCCCAGAAGACCTGCCTGATCTGT
ArgAspHisValLeuProIleAspTyrTyrPheProProGlnLysThrCysLeuIleCys

5 amino acid linker

FIGURE 6B
RIGHT

```
             170         180         190         200
ATG CAA ACA CAA AAA CCG ACT CTC GAA CTG CTA ACC TGC GAA GGC
MET GLN THR GLN LYS PRO THR LEU GLU LEU LEU THR CYS GLU GLY
 1                                   10
             260         270         280         290
GAT CGT CCG GCA ACG CTG CTG CTG GAA TCC GCA GAT ATC GAC AGC
ASP ARG PRO ALA THR LEU LEU LEU GLU SER ALA ASP ILE ASP SER
                                     40
             350         360         370         380
ATT ACA GCT TTA GGT GAC ACT GTC ACA ATC CAG GCA CTT TCC GGC
ILE THR ALA LEU GLY ASP THR VAL THR ILE GLN ALA LEU SER GLY
                                     70
             440         450         460         470
GGT GTG GAA AGT GAA CAA TCA CCA AAC TGC CGT GTG CTG CGC TTC
GLY VAL GLU SER GLU GLN SER PRO ASN CYS ARG VAL LEU ARG PHE
                                    100
             530         540         550         560
CTT TCG GTT TTT GAC GCT TTC CGT TTA TTG CAG AAT CTG TTG AAT
LEU SER VAL PHE ASP ALA PHE ARG LEU LEU GLN ASN LEU LEU ASN
                                    130
```

FIGURE 7A
TOP LEFT

```
210                 220                 230                 240                 250
GCT TAT CGC GAC AAT CCC ACC GCG CTT TTT CAC CAG TTG TGT GGG
ALA TYR ARG ASP ASN PRO THR ALA LEU PHE HIS GLN LEU CYS GLY
                     20                                                          30
300                 310                 320                 330                 340
AAA GAT GAT TTA AAA AGC CTG CTG CTG GTA GAC AGT GCG CTG CGC
LYS ASP ASP LEU LYS SER LEU LEU LEU VAL ASP SER ALA LEU ARG
                     50                                                          60
390                 400                 410                 420                 430
AAC GGC GAA GCC CTC CTG GCA CTA CTG GAT AAC GCC CTG CCT GCG
ASN GLY GLU ALA LEU LEU ALA LEU LEU ASP ASN ALA LEU PRO ALA
                     80                                                          90
480                 490                 500                 510                 520
CCC CCT GTC AGT CCA CTG CTG GAT GAA GAC GCC CGC TTA TGC TCC
PRO PRO VAL SER PRO LEU LEU ASP GLU ASP ALA ARG LEU CYS SER
                    110                                                         120
570                 500                 590                 600                 610
GTA CCG AAG GAA GAA CGA GAA GCC ATG TTC TTC AGC GGC CTG TTC
VAL PRO LYS GLU GLU ARG GLU ALA MET PHE PHE SER GLY LEU PHE
                    140                                                         150
```

FIGURE 7A
TOP RIGHT

```
        620             630               640             650
TCT TAT GAC CTT GTG GCG GGA TTT GAA GAT TTA CCG CAA CTG TCA
SER TYR ASP LEU VAL ALA GLY PHE GLU ASP LEU PRO GLN LEU SER
                                     160
        710             720               730             740
CTG ATG GTG ATT GAC CAT CAG AAA AAA AGC ACC CGT ATT CAG GCC
LEU MET VAL ILE ASP HIS GLN LYS LYS SER THR ARG ILE GLN ALA
                                     190
        800             810               820             830
CGC CTG AAC GAA CTA CGT CAG CAA CTG ACC GAA GCC GCG CCG CCG
ARG LEU ASN GLU LEU ARG GLN GLN LEU THR GLU ALA ALA PRO PRO
                                     220
        890             900               910             920
AGC GAT GAA GAG TTC GGT GGC GTA GTG CGT TTG TTG CAA AAA GCG
SER ASP GLU GLU PHE GLY GLY VAL VAL ARG LEU LEU GLN LYS ALA
                                     250
        980             990               1000            1010
TCT CTG CCC TGC CCG TCA CCG CTG GCG GCC TAT TAC GTG CTG AAA
SER LEU PRO CYS PRO SER PRO LEU ALA ALA TYR TYR VAL LEU LYS
                                     280
       1070            1080               1090            1100
TTC ACC CTA TTT GGC GCG TCG CCG GAA AGC TCG CTC AAG TAT GAT
PHE THR LEU PHE GLY ALA SER PRO GLU SER SER LEU LYS TYR ASP
                                     310
```

FIGURE 7A
LEFT BOTTOM

```
660                 670                 680                 690                 700
GCG GAA AAT AAC TGC CCT GAT TTC TGT TTT TAT CTC GCT GAA ACG
ALA GLU ASN ASN CYS PRO ASP PHE CYS PHE TYR LEU ALA GLU THR
                170                                                             180
750                 760                 770                 780                 790
AGC CTG TTT GCT CCG AAT GAA GAA GAA AAA CAA CGT CTC ACT GCT
SER LEU PHE ALA PRO ASN GLU GLU GLU LYS GLN ARG LEU THR ALA
                200                                                             210
840                 850                 860                 870                 880
CTG CCA GTG GTT TCC GTG CCG CAT ATG CGT TGT GAA TGT AAT CAG
LEU PRO VAL VAL SER VAL PRO HIS MET ARG CYS GLU CYS ASN GLN
                230                                                             240
930                 940                 950                 960                 970
ATT CGC GCT GGA GAA ATT TTC CAG GTG GTG CCA TCT CGC CGT TTC
ILE ARG ALA GLY GLU ILE PHE GLN VAL VAL PRO SER ARG ARG PHE
                260                                                             270
1020                1030                1040                1050                1060
AAG AGT AAT CCC AGC CCG TAC ATG TTT TTT ATG CAG GAT AAT GAT
LYS SER ASN PRO SER PRO TYR MET PHE PHE MET GLN ASP ASN ASP
                290                                                             300
1110                1120                1130
GCC ACC AGC CGC CAG ATT GAG ATC
ALA THR SER ARG GLN ILE GLU ILE
                320
```

FIGURE 7A
RIGHT BOTTOM

11 amino acid linker

```
CCC GGG CGA GCT CGA ATT CGA GCT CGC CCG GGG
Pro Gly Arg Ala Arg Ile Arg Ala Arg Pro Gly
```

```
                                                ATC TGT GGA
                                                ile cys gly
TGC AAG GTC TTC TTC AAA AGA GCT GCG GAA GGG AAA CAG AAG TAT
cys lys val phe phe lys arg ala ala glu gly lys gln lys tyr
AAA AAT TGT CCA TCG TGT CGT CTC CGG AAA TGT TAT GAA GCA GGG
lys asn cys pro ser cys arg leu arg lys cys tyr glu ala gly
CTA CAG GAA GAA GGA GAA AAC TCC AGT GCT GGT AGC CCC ACT GAG
leu gln glu glu gly glu asn ser ser ala gly ser pro thr glu
TGT CAA CCT ATC TTT CTT AAT GTC CTG GAA GCC ATT GAG CCA GGA
cys gln pro ile phe leu asn val leu glu ala ile glu pro gly
GCC TTG TTA TCT AGT CTC AAC GAG CTT GGC GAG AGA CAG CTT GTA
ala leu leu ser ser leu asn glu leu gly glu arg gln leu val
CAT GTG GAT GAC CAG ATG GCA GTC ATT CAG TAT TCC TGG ATG GGA
his val asp asp gln met ala val ile gln tyr ser trp met gly
TCT AGG ATG CTC TAC TTT GCA CCT GAC CTG GTT TTC AAT GAG TAT
ser arg met leu tyr phe ala pro asp leu val phe asn glu tyr
CAC CTT TCT CAA GAG TTT GGA TGG CTC CAG ATA ACC CCC CAG GAA
his leu ser gln glu phe gly trp leu gln ile thr pro gln glu
GAT GGG CTG AAA AAT CAA AAA TTC TTT GAT GAA CTT CGA ATG AAC
asp gly leu lys asn gln lys phe phe asp glu leu arg met asn

CCT CTA GAG TCG ACC TGC AGC CCA AGC TTA TCG ATG ATA AGC TGT CAA ACA TGA
Pro Leu Glu Ser Thr Cys Ser Pro Ser Leu Ser Met Ile Ser Cys Gln Thr ATOP
```

17 amino acid linker

Total amino acids: 323 + 11 + 279 + 17 = 630

FIGURE 7B
LEFT

```
GAT GAA GCT TCT GGT TGT CAC TAC GGA GCT CTC ACT TGT GGC AGC
asp glu ala ser gly cys his tyr gly ala leu thr cys gly ser 561

CTA TGT GCC AGC AGA AAT GAT TGC ACC ATT GAT AAA TTT CGG AGG
leu cys ala ser arg asn asp cys thr ile asp lys phe arg arg 591

ATG ACT CTG GGA GCT CGT AAG CTG AAG AAA CTT GGA AAT CTC AAA
met thr leu gly ala arg lys leu lys lys leu gly asn leu lys 621

GAC CCA TCC CAG AAG ATG ACT GTA TCA CAC ATT GAA GGC TAT GAA
asp pro ser gln lys met thr val ser his ile glu gly tyr glu 651

GTG GTG TGT GCC GGA CAT GAC AAC AAC CAG CCT GAT TCC TTT GCT
val val cys ala gly his asp asn asn gln pro asp ser phe ala 681

CAT GTG GTC AAG TGG GCC AAG GCC TTG CCT GGC TTC CGC AAC TTG
his val val lys trp ala lys ala leu pro gly phe arg asn leu 711

CTG ATG GTA TTT GCC ATG GGT TGG CGG TCC TTC ACT AAT GTC AAC
leu met val phe ala met gly trp arg ser phe thr asn val asn 741

CGC ATG CAC AAG TCT CGA ATG TAC AGC CAG TGC GTG AGG ATG AGG
arg met his lys ser arg met tyr ser gln cys val arg met arg 771

TTC CTG TGC ATG AAA GCA CTG CTA CTC TTC AGC ATT ATT CCA GTG
phe leu cys met lys ala leu leu leu phe ser ile ile pro val 801

TAC ATC AAG GAA CTT GAT
tyr ile lys glu leu asp
```

FIGURE 7B
RIGHT

```
        170         180         190         200
ATG CAA ACA CAA AAA CCG ACT CTC GAA CTG CTA ACC TGC GAA GGC
MET GLN THR GLN LYS PRO THR LEU GLU LEU LEU THR CYS GLU GLY
 1                                   10
        260         270         280         290
GAT CGT CCG GCA ACG CTG CTG CTG GAA TCC GCA GAT ATC GAC AGC
ASP ARG PRO ALA THR LEU LEU LEU GLU SER ALA ASP ILE ASP SER
                                     40
        350         360         370         380
ATT ACA GCT TTA GGT GAC ACT GTC ACA ATC CAG GCA CTT TCC GGC
ILE THR ALA LEU GLY ASP THR VAL THR ILE GLN ALA LEU SER GLY
                                     70
        440         450         460         470
GGT GTG GAA AGT GAA CAA TCA CCA AAC TGC CGT GTG CTG CGC TTC
GLY VAL GLU SER GLU GLN SER PRO ASN CYS ARG VAL LEU ARG PHE
                                    100
        530         540         550         560
CTT TCG GTT TTT GAC GCT TTC CGT TTA TTG CAG AAT CTG TTG AAT
LEU SER VAL PHE ASP ALA PHE ARG LEU LEU GLN ASN LEU LEU ASN
                                    130
        620         630         640         650
TCT TAT GAC CTT GTG GCG GGA TTT GAA GAT TTA CCG CAA CTG TCA
SER TYR ASP LEU VAL ALA GLY PHE GLU ASP LEU PRO GLN LEU SER
                                    160
```

FIGURE 8A
TOP LEFT

323 amino acids from TRP E protein

```
210             220             230             240             250
GCT TAT CGC GAC AAT CCC ACC GCG CTT TTT CAC CAG TTG TGT GGG
ALA TYR ARG ASP ASN PRO THR ALA LEU PHE HIS GLN LEU CYS GLY
                 20                                          30
300             310             320             330             340
AAA GAT GAT TTA AAA AGC CTG CTG CTG GTA GAC AGT GCG CTG CGC
LYS ASP ASP LEU LYS SER LEU LEU LEU VAL ASP SER ALA LEU ARG
                 50                                          60
390             400             410             420             430
AAC GGC GAA GCC CTC CTG GCA CTA CTG GAT AAC GCC CTG CCT GCG
ASN GLY GLU ALA LEU LEU ALA LEU LEU ASP ASN ALA LEU PRO ALA
                 80                                          90
480             490             500             510             520
CCC CCT GTC AGT CCA CTG CTG GAT GAA GAC GCC CGC TTA TGC TCC
PRO PRO VAL SER PRO LEU LEU ASP GLU ASP ALA ARG LEU CYS SER
                110                                         120
570             580             590             600             610
GTA CCG AAG GAA GAA CGA GAA GCC ATG TTC TTC AGC GGC CTG TTC
VAL PRO LYS GLU GLU ARG GLU ALA MET PHE PHE SER GLY LEU PHE
                140                                         150
660             670             680             690             700
GCG GAA AAT AAC TGC CCT GAT TTC TGT TTT TAT CTC GCT GAA ACG
ALA GLU ASN ASN CYS PRO ASP PHE CYS PHE TYR LEU ALA GLU THR
                170                                         180
```

FIGURE 8A
TOP RIGHT

```
       710                 720                 730                 740
CTG ATG GTG ATT GAC CAT CAG AAA AAA AGC ACC CGT ATT CAG GCC
LEU MET VAL ILE ASP HIS GLN LYS LYS SER THR ARG ILE GLN ALA
                                    190
       800                 810                 820                 830
CGC CTG AAC GAA CTA CGT CAG CAA CTG ACC GAA GCC GCG CCG CCG
ARG LEU ASN GLU LEU ARG GLN GLN LEU THR GLU ALA ALA PRO PRO
                                    220
       890                 900                 910                 920
AGC GAT GAA GAG TTC GGT GGC GTA GTG CGT TTG TTG CAA AAA GCG
SER ASP GLU GLU PHE GLY GLY VAL VAL ARG LEU LEU GLN LYS ALA
                                    250
       980                 990                1000                1010
TCT CTG CCC TGC CCG TCA CCG CTG GCG GCC TAT TAC GTG CTG AAA
SER LEU PRO CYS PRO SER PRO LEU ALA ALA TYR TYR VAL LEU LYS
                                    280
      1070                1080                1090                1100
TTC ACC CTA TTT GGC GCG TCG CCG GAA AGC TCG CTC AAG TAT GAT
PHE THR LEU PHE GLY ALA SER PRO GLU SER SER LEU LYS TYR ASP
                                    310
```

FIGURE 8A
BOTTOM LEFT

```
750                 760                 770                 780                 790
AGC CTG TTT GCT CCG AAT GAA GAA GAA AAA CAA CGT CTC ACT GCT
SER LEU PHE ALA PRO ASN GLU GLU GLU LYS GLN ARG LEU THR ALA
                200                                                             210
840                 850                 860                 870                 880
CTG CCA GTG GTT TCC GTG CCG CAT ATG CGT TGT GAA TGT AAT CAG
LEU PRO VAL VAL SER VAL PRO HIS MET ARG CYS GLU CYS ASN GLN
                230                                                             240
930                 940                 950                 960                 970
ATT CGC GCT GGA GAA ATT TTC CAG GTG GTG CCA TCT CGC CGT TTC
ILE ARG ALA GLY GLU ILE PHE GLN VAL VAL PRO SER ARG ARG PHE
                260                                                             270
1020                1030                1040                1050                1060
AAG AGT AAT CCC AGC CCG TAC ATG TTT TTT ATG CAG GAT AAT GAT
LYS SER ASN PRO SER PRO TYR MET PHE PHE MET GLN ASP ASN ASP
                290                                                             300
1110                1120                1130
GCC ACC AGC CGC CAG ATT GAG ATC
ALA THR SER ARG GLN ILE GLU ILE
                320
```

CCC CCG
Pro Pro

FIGURE 8A
BOTTOM RIGHT

```
                                                                    GAA TTC
                                                                    glu phe
GAT GGG CTG AAA AAT CAA AAA TTC TTT GAT GAA CTT CGA ATG AAC TAC
asp gly leu lys asn gln lys phe phe asp glu leu arg met asn tyr
CCC ACA TCC TGC TCA AGG CGC TTC TAC CAG CTC ACC AAG CTC CTG GAT
pro thr ser cys ser arg arg phe tyr gln leu thr lys leu leu asp
GAC CTG CTA ATC AAG TCC CAT ATG GTG AGC GTG GAC TTT CCT GAA ATG
asp leu leu ile lys ser his met val ser val asp phe pro glu met
GGG AAA GTC AAG CCC ATC TAT TTC CAC ACA CAG TGA
gly lys val lys pro ile tyr phe his thr gln ***
```

Total amino acids: 323 + 2 + 117 = 442

FIGURE 8B
TOP LEFT

2 amino acid linker

```
CTG TGC ATG AAA GCA CTG CTA CTC TTC AGC ATT ATT CCA GTG
leu cys met lys ala leu leu leu phe ser ile ile pro val 801

ATC AAG GAA CTT GAT CGC ATC ATT GCA TGC AAA AGA AAA AAT
ile lys glu leu asp arg ile ile ala cys lys arg lys asn 831

TCT GTG CAG CCT ATT GCA AGA GAG CTG CAT CAA TTC ACT TTT
ser val gln pro ile ala arg glu leu his gln phe thr phe 861

ATG GCA GAG ATC ATC TCT GTG CAA GTG CCC AAG ATC CTT TCT
met ala glu ile ile ser val gln val pro lys ile leu ser 891
```

FIGURE 8B
TOP RIGHT

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US89/01229

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) [6]

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC(4): C07H 21/04, C12N 1/20, C12N 15/00, C07K 13/00
U.S. Cl.: 536/27, 435/252.3,240.2,320,6,7,530/350,387

II. FIELDS SEARCHED

Minimum Documentation Searched [7]

Classification System	Classification Symbols
U.S.	536/27 435/68, 172.3, 240.2, 252.3, 320.6,

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched [8]

DATABASES: CHEMICAL ABSTRACTS ONLINE (FILE CA, 1967-1989; FILE BIOSIS, 1969-1989), USPTO AUTOMATED PATENT SYSTEM (FILE USPAT, 1975-1988). SEE ATTACHMENT FOR SEARCH TERMS.

III. DOCUMENTS CONSIDERED TO BE RELEVANT [9]

Category *	Citation of Document, [11] with indication, where appropriate, of the relevant passages [12]	Relevant to Claim No. [13]
X,P Y	Science (Washington, USA), Volume 240, Issued April 1988, Lubahn et al., "Cloning of human androgen receptor complementary DNA and localization to the X chromosome", pages 327-330, see the entire document.	1-3, 5, 10-13 6-8
X,P Y	Science (Washington, USA), Volume 240, Issued April 1988, Chang et al., "Molecular cloning of human and rat complementary DNA encoding androgen receptor", pages 324-326, see the entire document	1-3, 5,8, 10-13 6,7

* Special categories of cited documents: [10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
17 June 1989	24 JUL 1989
International Searching Authority	**Signature of Authorized Officer**
ISA/US	Jasemine C. Chambers JASEMINE C. CHAMBERS

FURTHER INFORMATION CONTINUED FROM THE SECOND SHEET

V. ☐ OBSERVATIONS WHERE CERTAIN CLAIMS WERE FOUND UNSEARCHABLE [1]

This international search report has not been established in respect of certain claims under Article 17(2) (a) for the following reasons:

1. ☐ Claim numbers , because they relate to subject matter [12] not required to be searched by this Authority, namely:

2. ☐ Claim numbers , because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out [13], specifically:

3. ☐ Claim numbers___________, because they are dependent claims not drafted in accordance with the second and third sentences of PCT Rule 6.4(a).

VI. ☒ OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING [2]

This International Searching Authority found multiple inventions in this international application as follows:

I. Claims 1-3, 5-8, 10-13, 30 and 32, drawn to androgen receptor DNA, plasmid, cell and method of use of DNA; Clas 435, subclasses 6, 240.2, 252.3 and 320, and Class 536, subclass 27. See attachment.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims of the international application.

2. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims of the international application for which fees were paid, specifically claims:

3. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claim numbers:
1-3, 5-8, 10-13, 30 and 32. Telephone practice.

4. ☐ As all searchable claims could be searched without effort justifying an additional fee, the International Searching Authority did not invite payment of any additional fee.

Remark on Protest

☐ The additional search fees were accompanied by applicant's protest.

☐ No protest accompanied the payment of additional search fees.

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No
Y	Cold Spring Harbor Symposia on Quantitative Biology, Cold Spring Harbor Laboratory (New York, USA), Volume LI, Published 1986, Mullis et al.," Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction", pages 263-273, see the entire document.	30,32
Y	Nature (London, UK), Volume 324, Issued November 1986, Saiki et al., "Analysis of enzymatically amplified B-globin and HLA-DQ DNA with allele-specific oligonucleotide probes", pages 163-166, see the entire document.	30, 32
Y,P	US, A, 4,800,159 (MULLIS et al.) 24 JANUARY 1989, see the entire document.	30,32
X,P	Biochemical and Biophysical Research Communications, Academic Press (Orlando, USA), Volume 153, Issued May 1988, Trapman et al., "Cloning, Structure and expression of a cDNA encoding the human androgen receptor", pages 241-248, see the entire document.	1-3, 5,8, 10-13 / 6,7
X,P / Y	Proceedings ofthe National Academy of Sciences, USA (Washington, USA), volume 85, Issued October 1988, Chang et al., "Structural analysis of complementary DNA and amino acid sequences of human and rat androgen receptors", pages 7211-7215, see the entire document.	1-3, 5,8 10-13 / 6,7
X / Y	Journal of Endocrinological Investigation (Milan, Italy), Volume 10, Supplement 2, Published 1987, Govindan et al., "Cloning of the human androgen receptor cDNA", page 63, see the entire abstract.	1-3, 5, 10-13 / 6-8
X,P / Y	Progress in Cancer Research and Therapy, Raven Press (New York, USA), Volume 35, Issued July 1988, Govindan et al., "Cloning of the human androgen receptor cDNA", pages 49-54, see the entire document.	1-3, 5, 10-13 / 6-8